(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2001 (26.04.2001)

PCT

(10) International Publication Number
WO 01/29242 A2

(51) International Patent Classification[7]: C12N 15/82, A01H 5/00

(21) International Application Number: PCT/US00/29027

(22) International Filing Date: 20 October 2000 (20.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/160,758 21 October 1999 (21.10.1999) US
60/195,282 7 April 2000 (07.04.2000) US

(71) Applicant: MONSANTO COMPANY [US/US]; 800 North Lindbergh Blvd., Saint Louis, MO 63167 (US).

(72) Inventors: RUSSELL, Douglas; 1610 Frisch Road, Madison, WI 53711 (US). MANJUNATH, Siva; 1372 Whispering Pines Drive, Saint Louis, MO 63146 (US). BASSUNER, Ronald; 5079 Sunrise Ridge Trail, Middleton, WI 53562 (US).

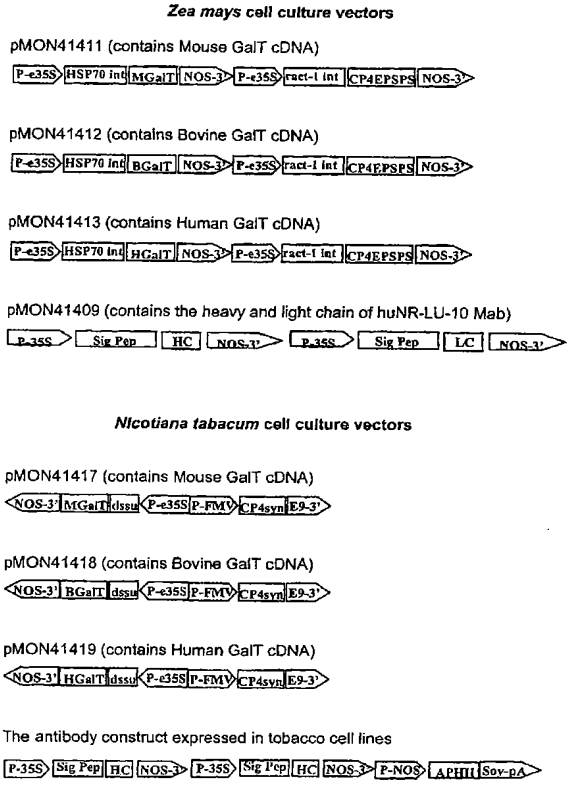
(74) Agents: PELTO, Don et al.; McKenna & Cuneo, LLP, 1900 K Street, NW, Washington, DC 20006 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,

[Continued on next page]

(54) Title: POST-TRANSLATIONAL MODIFICATION OF RECOMBINANT PROTEINS PRODUCED IN PLANTS



(57) Abstract: The present invention relates to a process for the production of proteins or polypeptides using genetically manipulated plants or plant cells, as well as to the genetically manipulated plants and plant cells per se (including parts of the genetically manipulated plants), the heterologous protein material (e.g., a protein, polypeptide and the like) which is produced with the aid of these genetically manipulated plants or plant cells, and the recombinant polynucleotides (DNA or RNA) that are used for the genetic manipulation.

WO 01/29242 A2

IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

POST-TRANSLATIONAL MODIFICATION OF RECOMBINANT PROTEINS PRODUCED IN PLANTS

CROSS REFERENCE TO RELATED APPLICATIONS

The present invention is related to and claims the benefit of, under 35 U.S.C. § 120, provisional patent applications Serial No. 60/160,758, filed 20 October, 1999, and Serial No. 60/195,282, filed 07 April, 2000, both of which are expressly incorporated fully herein by reference.

FIELD OF INVENTION

The present invention is in the field of recombinant DNA technology, more in particular recombinant DNA technology concerning the genetic manipulation of plants, and is related to a process for the production of proteins or polypeptides using genetically manipulated plants or plant cells, as well as to the genetically manipulated plants and plant cells per se (including parts of the genetically manipulated plants), the heterologous protein material (e.g., a protein, polypeptide and the like) which is produced with the aid of these genetically manipulated plants or plant cells, and the recombinant polynucleotides (DNA or RNA) and associated vectors that are used for the genetic manipulation.

BACKGROUND OF THE INVENTION

Recombinant DNA technology has enabled the production of foreign, recombinant polypeptides in host systems. Indeed, expression vectors have been constructed for highly efficient expression of foreign proteins in numerous host systems. These expression vectors contain a wide variety of structural genes, isolated from mammals and viruses, which are operably linked to diverse regulatory regions functional in the host system. The expression vector may be introduced into the host system, and the transformed host system may then be cultured and subjected to conditions which facilitate the expression of the structural gene and lead to the expression of large quantities of the desired protein. Much of the early work in biotechnology was directed toward the expression of recombinant or "heterologous" proteins in prokaryote hosts, such as *Escherichia coli* and *Bacillus subtilis*. Such work in prokaryote hosts provided ease of genetic manipulation, rapid growth of the organisms in batch culture, and the possibility of large-scale fermentation. Examples of heterologous proteins include

insulin, various interferon types, and human growth hormone which are produced in commercial quantities by, for example, bacteria and yeast cells.

The advantages of producing a recombinant protein from a prokaryote host in lieu of isolation from a natural source include: the ready availability of raw material; high expression levels, which is especially useful for proteins of low natural abundance; the ease with which a normally intracellular protein can be excreted into the expression medium, facilitating the purification process; and the relative ease with which modified (fusion) proteins can be created to further simplify the purification of the resultant protein.

Major disadvantages, however, accompany these advantages. Importantly, the required elements of the active protein which result from post-translational modification may not be carried out in the prokaryote host. Some of these post-translational modifications include signal peptide processing, pro-peptide processing, protein folding, disulfide bond formation, glycosylation, gamma carboxylation, and beta-hydroxylation. As a result, complex proteins derived from prokaryote hosts are not always properly folded or processed to provide the desired degree of biological activity. Consequently, prokaryote hosts have generally been utilized for the expression of relatively simple foreign polypeptides that do not require folding or post-translational processing to achieve a biologically active protein. Indeed, the costs associated with the inability of bacteria to perform many of the post-translational modifications required for the biological activity of recombinant proteins of mammals limit the value of this host system. More specifically, extensive post-purification chemical and enzymatic treatments can be required to obtain biologically active protein

Due to the disadvantages associated with prokaryotic hosts, the biotechnology industry had directed its efforts to eukaryotic hosts like mammalian cell tissue culture, yeast, fungi, insect cells, and transgenic animals, to properly and efficiently express recombinant proteins. These hosts, however, may suffer from any or all of the following disadvantages: expensive fermentation, low yields, secretion problems, inappropriate modifications in protein processing, high operating costs, difficulties in scaling up to large volumes, contamination that kills the host culture, and potential contamination by virus or prion pathogens. For these reasons, existing eukaryotic hosts are unable to provide high-volume, low-cost protein production of heterologous proteins.

Moreover, although mammalian cells are capable of correctly folding and glycosylating bioactive proteins, the quality and extent of glycosylation can vary with

different culture conditions among the same host cells. Yeast, alternatively, produce incorrectly glycosylated proteins that have excessive mannose residues, and generally exhibit limited post-translational processing. Other fungi may be available for high-volume, low-cost production, but they are not capable of expressing many target proteins. As another example, the baculovirus insect cell system can produce high levels of glycosylated proteins: these proteins are not secreted, however, thus making purification complex and expensive. Transgenic animals are subject to lengthy lead times to develop herds with stable genetics, high operating costs, and contamination by prions or viruses.

The biochemical, technical and economic limitations on existing prokaryotic and eukaryotic expression systems has created substantial interest in developing new expression systems for the production of heterologous proteins. To that end, plants represent a suitable alternative to other host systems because of the advantageous economics of growing plant crops, plant suspension cells, and tissues such as callus; the ability to synthesize proteins in storage organs like tubers, seeds, fruits and leaves; and the ability of plants to perform many of the post-translational modifications previously described. Strum et al., 175 PLANTA 170-83 (1988).

Therefore, it is desirable to produce heterologous proteins from a source such as plants, which offer the opportunity for the "Molecular Farming" of important proteins. *See, e.g.*, U.S. Pat. No. 5,550,038, which is expressly incorporated herein by reference. Transgenic plants have been studied over the past several years for potential use in low cost production of high quality, biologically active mammalian proteins. *See, e.g.*, Sijmons et al., 8 BIO/TECHNOLOGY 217-21 (1990); Vandekerckhove et al., 7 BIO/TECHNOLOGY 929-32 (1989); Conrad & Fiedler, 26 PLANT MOLEC. BIOL. 1023-30 (1994); Ma et al., 268 SCIENCE 716-19 (1995). Expression of monoclonal antibodies in plant host systems has been widely studied primarily due to their potential value as therapeutic and clinical reagents.

For example, the synthesis of functional IgG antibodies has been reported in plant host systems. *See* Düring, Inaugural Dissertation (1988); Düring & Hippe, 370 BIOL. CHEM. HOPPE SEYLER 888 (1989); Düring et al., 15 PLANT MOL. BIOL. 281-93 (1990). These plant host systems include *Nicotania tabacum* (tobacco) plants, capable of expressing IgG antibodies. Hiatt et al., 342 NATURE 76-78 (1989); Ma et al., 24 EUR. J. IMMUNOL. 131-38 (1994); U.S. Pat. Nos. 5,202,422 and 5,639,947. More recently, a more complex IgA antibody was synthesized in transgenic tobacco plants. U.S. Pat. No. 5,959,177. The

synthesis of IgA in rice has been reported recently as well. WO 99/66,026. Antibodies expressed in *Zea mays* (corn) plants include monoclonal antibody BR96 and monoclonal antibody NeoRx451 (WO 98/10,062).

Single-chain antibody fragments are well-known in the art. Bird et al., 242 SCIENCE 423-26 (1988). Functional single chain fragments have been successfully expressed in the leaves of tobacco and *Arabidopsis* plants. Owen et al. 10 Bio/Tech. 790-94 (1992); Artsaenko et al., 8 PLANT J. 745-50 (1995); Fecker et al., 32 PLANT MOL. BIOL. 979-86 (1996). Long term storage of single chain antibody fragments has also been indicated in tobacco seeds. Fielder et al. 13 BIO/TECH. 1090-93 (1995). L6 sFv single chain anti-carcinoma antibody, anti-TAC sFv (that recognizes L2 receptor) and G28.5 sFv single-chain antibody (that recognizes CD40 cell surface protein) have been produced in high levels in tobacco culture. U.S. Pat. No. 6,080,560. Additionally, the single-chain antibody L6 has been successfully produced in corn and soy. Cooley et al., 108(2) PLANT PHYSIOL. 150 (1995).

Other heterologous proteins expressed in plants include human serum albumin (HSA), secreted into the medium from plant cells derived from both potato and tobacco plants. Sijmons et al., 8 BIO/TECHNOLOGY 217-21 (1990). Additionally, various other proteins have been successfully produced in plants. *See, e.g.,* Kusnadi et al., 56(5) BIOTECH. & BIOENG. 473-84 (1997); U.S. Pat. No. 5,550,038. Human serum albumin, transgenic plant rabbit liver cytochrome P450, hamster 3-hydroxy-3-methylglutaryl CoA reductase, and the hepatitis B surface antigen have been reported in the art. *See, e.g.,* Sijmons et al., 8 BIO/TECH. 217-21 (1990); Saito et al., 88 P.N.A.S. 7041-45 (1991); Mason et al., 89 P.N.A.S. 11745-49 (1992).

The heterologous polypeptide of this invention is preferably a eukaryotic, non-plant protein, most preferably of mammalian origin. Preferably, the heterologous peptide either requires post-translational modification or is, itself, a post-translational modifying enzyme. The heterologous polypeptide requiring post-translational modification may be an antibody, a fragment of an antibody, collagen, protein C, erythropoietin, other therapeutic molecules or blood substitutes, cytokines, or other proteins with pharmaceutical or nutritional value.

Other objectives, features and advantages of the present invention will become apparent from the following detailed description. The detailed description and the specific examples, while indicating specific embodiments of the invention, are provided by way of illustration only. Accordingly, the present invention also includes those various changes and

modifications within the spirit and scope of the invention that may become apparent to those skilled in the art from this detailed description.

SUMMARY OF THE INVENTION

The present invention is directed to methods for producing a post-translationally modified heterologous polypeptide in a plant host system by altering the natural post-translational modification abilities of that plant host system. In a preferred embodiment, this method includes transforming a plant host system with a nucleic acid that encodes a heterologous polypeptide, and isolating that polypeptide from the plant host system.

In another preferred method, the post-translational modification may be proteolytic cleavage, glycosylation, phosphorylation, methylation, sulfation, prenylation, acetylation, N-amidation, oxydation, hydroxylation, or myristylation. In another aspect of this method, altering the natural post-translational modifications is done by transforming the plant host system with one or more nucleic acid sequences encoding a post-translational modification enzyme. In an alternative aspect, the altering is done by mutagenesis of plant host system. In another embodiment, the altering is done by transforming said plant host system with an expression vector comprising a nucleic acid sequence that encodes an antisense nucleic acid.

In another aspect of the present invention, the post-translationally modified enzyme is a heterologous polypeptide such as immunoglobulins, integrins, addressins, selectins, homing receptors, T-cell receptor units, soluble major histocompatibility complex antigens, growth factor receptors, growth factors, growth hormones, cell cycle proteins, viral antigens, bacterial antigens, vaccines, fibrinogen, thrombin and hyaluronic acid.

In a preferred embodiment of the invention, the plant host system is a plant.

This invention also related to a method for producing a post-translationally modified heterologous polypeptide in a plant host system comprising the step of expressing said heterologous polypeptide, wherein the cells of said plant host system have been transformed with one or more expression vectors comprising a nucleic acid sequence encoding a heterologous polypeptide.

The instant invention also provides a method for producing a post-translationally modified heterologous polypeptide by expressing said post-translational modifying enzyme, in the cells of the plant host system hat have been transformed with an expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

Another embodiment of the present invention provides a method for producing a post-translationally modified heterologous polypeptide, wherein the cells of the plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme, by expressing the heterologous polypeptide and expressing said post-translational modifying enzyme.

The instant invention further provide a plant host system expressing a post-translationally modified heterologous polypeptide, wherein the natural post-translational modification abilities of said plant host system have been altered.

Another embodiment of the invention is a plant host system that expresses a post-translationally modified heterologous polypeptide, wherein the cells of the plant system have been transformed with an expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide.

An alternative embodiment is a plant host system that expresses a post-translationally modified heterologous polypeptide, wherein the cells of said plant host system have been transformed with an expression vector comprising a post-translational modifying enzyme.

Another aspect of the invention is a plant host system expressing a post-translationally modified heterologous polypeptide, wherein the cells of the plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

The invention described herein further provides a method for producing a post-translationally modified heterologous polypeptide in a plant host system, by cross-pollinating a first plant, wherein the plant has been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide, and a second plant wherein the second plant has been transformed with a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

Another aspect of the invention provides a plant host system that produces a post-translationally modified heterologous polypeptide wherein the plant host system expresses a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second express vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

The present invention relates to the production of post-translationally modified heterologous proteins comprising the expression of a heterologous protein gene, and at least one set of genes encoding a mammalian-specific post-translational enzyme in a plant host system. The heterologous proteins may include antibodies and antibody fragments, collagen types I-XX, human protein C, and cytokines.

More specifically, the present invention provides for methods of expressing at least one nucleic acid molecule encoding a heterologous protein within a plant host system or plant cell. The plant host system may be any monocotyledonous or dicotyledonous plant. The monocotyledonous plants include, but are not limited to, corn, cereals, grains, grasses, and rice. The dicotyledonous plants may include, but are not limited to, tobacco, tomatoes, potatoes, and legumes including soybean and alfalfa. The present invention further provides for expression of more than one nucleic acid molecule encoding a heterologous protein within a plant host system or plant cell, so that a number of different genes are introduced and expressed. Furthermore, the present invention provides methods of further introducing and expressing in a plant host system or plant cell at least one nucleic acid sequence encoding mammalian-specific post-translational modifying enzymes involved in the post-translational modification of the expressed heterologous protein. The nucleic acid sequence encoding the heterologous protein, fragments, variants, mutants or hybrids of such nucleic acid sequence, and the nucleic acid sequences encoding enzymes involved in the post-translational modification of fragments, variants, mutants or hybrids of such nucleic acid sequence, are contained within the plant host system or plant cell to produce at least one heterologous protein or fragment or functional equivalents thereof.

In another aspect, the present invention provides a plant host system expressing a post-translational enzyme or subunit thereof important to the biosynthesis of mammalian proteins. In yet another aspect, the present invention provides a plant host system transformed with a nucleic acid sequence encoding a heterologous protein, which may be combined with other factors that drive or control the expression of the protein or functional equivalents thereof. Such factors include transcriptional promoters that are operably linked to a coding sequence of the protein or functional equivalent thereof, or a post-translational enzyme.

Also within the scope of the present invention are expression vectors comprising of a polynucleotide sequence encoding the desired heterologous protein, variants, mutants or hybrids of such nucleic acid sequence, necessary promoters, and other sequences important

for the proper expression of heterologous protein in transgenic plants. The vectors and plasmids of the present invention further may comprise at least one nucleic acid sequence, fragment, variant, hybrid, or mutant thereof encoding one or more heterologous protein or post-translational enzymes or functional equivalents thereof.

Plant cells may not naturally produce sufficient amounts of post-translational enzymes to efficiently produce functional and/or stable mammalian heterologous proteins. In one aspect of the present invention, such plant cells are transformed with expression vectors comprising nucleic acid sequences encoding post-translational enzymes. The present invention provides plant cells that are transformed and thus supplemented with the necessary post-translational enzymes to produce functional and/or stable proteins. In a specific embodiment of the present invention, the post-translational enzyme may be β-(1,4)-galactosyltransferase, wherein the nucleic acid sequence encoding β-(1,4)-galactosyltransferase is introduced into plant host systems or plant cells to produce a biologically active β-(1,4)-galactosyltransferase. In an alternate embodiment of the present invention, the post-translational enzyme may be prolyl 4-hydroxylase, wherein the nucleic acid sequences encoding an α subunit of prolyl 4-hydroxylase and the nucleic acid sequences encoding a β subunit of prolyl 4-hydroxylase are transformed into plant host systems or plant cells to produce a biologically active prolyl 4-hydroxylase enzyme. In a further embodiment of the present invention, the post-translation enzyme may be γ-glutamyl carboxylase, wherein the nucleic acid sequences encoding γ-glutamyl carboxylase are transformed into plant host systems or plant cells to produce a biologically active γ-glutamyl carboxylase enzyme.

In another aspect of the present invention, the plant host system or plant cell may be genetically manipulated, thereby altering the natural post-translational modifying abilities of the plant host system or plant cell. The plant host system may be genetically manipulated by transforming the plant host system with an expression vector that encodes at least one mammalian-specific post-translational modification enzyme. The mammalian-specific post-translational modification enzymes include, but are not limited to, prolyl 4-hydroxylase or a subunit thereof, lysyl oxidase, lysyl hydroxylase, C-proteinase, N-proteinase, PACE, γ-glutamyl carboxylase, N-acetylglucosaminaltransferases, N-acetlygalactosaminyltransferases, N-acetlygalactosaminyltransferases, sialyl-transferases, fucosyltransferases, galactosyltransferases, mannosyltransferases, sulfotransferases, glycosidases,

acetyltransferases, and mannosidases. An alternative methodology for altering the natural post-translational modifying abilities of a plant host system would comprise transforming the plant host system with at least one nucleic acid sequence that expresses at least one enzyme capable of cleaving plant-specific linkages. Plant specific linkages would include, but are not limited to, β(1,2)-xylose and α(1,3)-fucose. In a further aspect, the natural post-translational modification abilities of the plant host system may be altered by transforming the plant host system with a nucleic acid that blocks protein production of one or more plant specific post-translational modifying enzymes. The nucleic acid may be a DNA or RNA molecule that is complementary to a plant-specific post-translational modifying enzyme. These nucleic acids would include antisense DNA and antisense RNA. Such plant specific post-translational modifying enzymes include, but are not limited to, galactosyl transferase, xylosyl transferase, and fucosyl transferase.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows a subset of mammalian asparagine-linked glycosylation patterns from antibodies. Sugars in italics are not always present in mammalian systems. The mammalian R groups can be one of the following: (a) R=GlcNac β(1, 2); (b) R=Gal β(1, 4)-GlcNac β(1, 2); (c) R=NeuAc α(2, 3)-Gal β(1, 4)-GlcNac β(1, 2); (d) R=NeuGc α(2, 3)-Gal β(1, 4)-GlcNac β(1, 2); and (e) R=Gal α(1, 3)-Gal β(1, 4)-GlcNac β(1, 2).

Figure 2 illustrates a subset of plant asparagine-linked glycosylation patterns from antibodies. Sugars in italics are not always present in plant systems. The plant R groups can be one of the following: (a) R=null; (b) R=GlcNac β(1, 2);
(c) R=Gal β (1,3)-GlcNac β (1,2) and (d) R=Gal β (1,4)-GlcNac β (1,2)

Fuc α (1,4)	Fuc α (1,6).

Figure 3 illustrates the results of human trials with huNR-LU-10, non-glycosylated humanized antibody produced from corn, compared with the glycosylation-modified murine antibody. The results indicate similar blood clearance kinetics both before and after clearing agent administration.

Figure 4 shows the design of primers for amplifying mouse, bovine, and human GalT nucleic acids using RT-PCR techniques.

Figure 5 shows RT-PCR amplification of mouse, bovine, and human GalT. Poly A+ RNA from mouse, bovine, and human liver tissues (from Clontech Laboratories, Inc., Palto Alto, CA) were used with gene-specific primers (as shown in Figure 4 – Lanes 1 through 8) to amplify mouse, bovine, and human GalT cDNA as described.

Figure 6 illustrates the constructs designed for the expression of mouse, bovine, and human GalT in *Zea mays* and *Nicotiana tabacum* cell cultures.

Figure 7 illustrates RT-PCR analysis of transformed *N. t.*-GalT calli demonstrating the accumulation of mRNA corresponding to mouse, bovine and human GalT in plant cells. Total RNA from GalT/MAB transgenic calli were extracted using Trizol reagent method (Life Technologies, Rockville, MD). Approximately 1.0 μg total RNA was used in the RT-PCR reaction along with gene-specific primers. Aliqouts (10 μl) from each PCR reaction were separated on an agarose gel. Odd-numbered lanes correspond to the reaction samples that contain reverse transcriptase enzyme and the even-numbered lanes correspond to control samples without reverse transcriptase. Lanes 21 and 22: Demonstrate the expression of heavy chain mRNA in transgenic samples that also show the expression human GalT mRNA (lane 19); Lane 23: Control reaction sample without primer; Lane 24: Control reaction sample without mRNA template.

Figure 8 illustrates the positive transformation events identified through Northern blot analysis of corn GalT transgenic calli. ** indicates a sample (2D1-2) that gave a strong signal, and served as positive controls in other blots. * indicates that the signal in these samples is similar to the signal obtained for sample 2D1-2. The mouse GalT transgenic samples begin with the numeral 2. Fourteen out of 24 screened showed mRNA accumulation corresponding to the mouse GalT. The bovine GalT samples begin with the numeral 3. Eight out of 24 samples screened showed mRNA accumulation corresponding to the bovine GalT. The human GalT samples begin with the numeral 4. Eleven out of 24 samples showed mRNA accumulation corresponding to human GalT. Samples 2D1-2, 2B2-4, 4A2-3, and 4B1-4 were obtained from 100 ml nutrient media.

Figure 9 illustrates the positive transformation events identified through Northern blot analysis of tobacco transgenic calli that have been transformed with either mouse, bovine, or human GalT.

Figure 10 depicts the expression of the huNR-LU-10 (NRX) antibody in transgenic corn GalT calli. TSP is the Total Soluble Protein.

Figure 11 shows the galactosyltransferase activity in transgenic corn callus. The samples beginning with the numeral 2 were transformed with mouse GalT. The samples beginning with the numeral 3 were transformed with bovine GalT. The sample designated 418-30 refers to tobacco cells also transformed with bovine GalT.

Figure 12 summarizes the screening of corn transgenic calli for the expression of GalT activity. The mouse GalT samples begin with the numeral 2. The bovine GalT samples begin with the numeral 3. The human GalT samples begin with the numeral 4.

Figure 13 depicts the expression cassette for endosperm-specific expression of bovine GalT in corn seed.

Figure 14 shows the alignment of DNAs encoding various glycosyltransferases and hybrid GalT molecules.

Figure 15 depicts the expression vectors designed for endosperm-specific expression of hybrid GalT enzymes in corn seed.

Figure 16 illustrates the glycosyltransferase biosynthetic pathway in plants. "X"s identify particular enzymes that may be targeted to alter the plant biosynthetic pathway and the ultimate glycosylation pattern of the target polypeptide.

Figure 17 depicts the expression vectors constructed for endosperm-specific expression of anti-glycosyltransferase antisense molecules in corn seed.

Figure 18 is a summary of the plant expression vectors used in the experiments related to the production of collagen in plants.

Figure 19 illustrates a Northern blot analysis. Plant cell culture samples were examined by Northern blot analysis to determine if the collagen cassette had successfully integrated to allow transcription of the full-length RNA. The probe was a 958 bp fragment of human collagen type III isolated from pwrg4710 and randomly labeled with P^{32}. The samples with suitable levels and size of hybridization signal are underlined.

Figure 20 shows a Western blot analysis. Plant suspension cells were analyzed by Western blot for prolyl 4-hydroxylase α, prolyl 4-hydroxylase β, and collagen. Lanes 1-5 represent the pwrg4723/4724 cell cultures; lanes 6-7 represent the pwrg4715/4716 cultures; lanes 8-9 represent the pwrg4723/4716 cultures. Lane 10 is null, and lanes 11-13 represent a dilution series of purified control standards of these proteins.

Figure 21 illustrates a Western blot analysis. Plant suspension cells (top panel) were subjected to analysis by Western blot for prolyl 4-hydroxylase β and collagen, and were

compared to the media from the same culture (bottom panel). Lanes 1-6 represent pwrg4723/4724 transformant cultures; lanes 7-12 represent pwrg4715/4716 transformant cultures; lane 13 represents a pwrg4723/4716 transformant culture and lane 14 represents the protein standards.

Figure 22 represents the deduced amino acid sequences at the ligation point of the signal peptides for prolyl 4-hydroxylase α, prolyl 4-hydroxylase β, and collagen.

Figure 23 illustrates the expression of prolyl 4-hydroxylase α, prolyl 4-hydroxylase β, and collagen in plant cells utilizing either the native signal peptide or *Nicotiana plumbaginofolia* extensin 5' UTR and signal peptide.

Figure 24 shows a Western blot of total soluble protein extracts of potato tubers expressing the huNR-LU-10 monoclonal antibody. Lanes 25, 26, and 29 through 36 refer to tuber samples of individual transgenic plants. Lanes 27 and 28 are non-transgenic controls. Tx, X, and Y are purified antibody standards. Asterisks indicate fully assembled antibodies.

DETAILED DESCRIPTION OF THE INVENTION

It is understood that the present invention is not limited to the particular methodology, protocols, cell lines, vectors, and reagents, *etc.*, described herein, as these may vary. It is also to be understood that the terminology used herein is used for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention. It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "an antibody" is a reference to one or more antibodies and includes equivalents thereof known to those skilled in the art and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Preferred methods, devices, and materials are described, although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention. All references cited herein are incorporated by reference herein in their entirety.

In a broad sense, the invention provides various methods for the production of a heterologous protein in a plant host system by transforming the plant host system with one or more expression vectors. The expression vectors may contain at least one structural

component, such as the heterologous protein and/or a post-translational modifying enzyme, and the appropriate regulatory sequences, such as promoters, transcriptional and translational initiation and termination signals, and signal sequences that are functional in the plant host system.

In a broad sense, the invention thus provides a process for the production of a heterologous polypeptide or protein in a plant host system, by growing plants or plant cells that, by means of genetic manipulation, are capable of effecting desired post-translational modifications that result in a final desired polypeptide or protein product that may not otherwise be so-modified in plant host systems. For example, one embodiment of the invention provides for the production desired in a plant host by using a recombinant nucleic acid that provides the plant host with the means of expressing the heterologous polypeptide or protein, and contains at least a nucleic acid sequence encoding the mature protein material, and/or a nucleic acid sequence encoding a post-translational modifying enzyme.

Thus, in one embodiment, this requires the introduction into the plants not only of the nucleic acid sequence or sequences encoding the structural peptide of interest, but the introduction of the nucleic acid sequence or sequences encoding post-translational enzymes or other molecules capable of modifying the peptide of interest or capable of altering the natural post-translation modification abilities of the host.. Such a molecule could cause the desired characteristic, or reverse an undesirable modification, or block the function of native enzymes, or be a signal peptide or ligand that acts to sequester the target protein in a particular sub-cellular compartment thus blocking undesirable maturation.

Other methods of attaining a plant host with the ability to provide the desired post-translational modification include classic mutant screening for the desired characteristic with or without prior mutagenesis, such as the elimination of a cleavage enzyme's activity (protease, glycosidase, deglycosidase, etc.) or the over-expression of cleavage enzymes that reverse the activity of conjugating enzymes. Plant-specific enzymes may include, but are not limited to, galactosyl transferase, xylosyl transferase, and fucosyl transferase. Another approach involves isolating a native enzyme with the desired activity and altering its expression to achieve desirable levels of activity by, e.g., linking it to a different transcription or translation regulatory element. Antisense targeting against an undesirable pathway, e.g., halting translation of an undesirable native enzyme, is another alternative. Alternatively, site directed mutagenesis may disable an undesirable gene at the chromosomal level (knock-out

technology). *See, e.g.*, U.S. Pat. No. 5,871,984. Alternatively, affecting retention of a target protein during synthesis may be desirable. For example, retention of a glycoprotein in the early pathway of glycogenesis to limit addition of particular residues, is possible. More specifically, this could be achieved by the direct fusion of a desired recombinant glycoprotein to an endoplasmic reticulum (ER) retention signal, or engineering a binding function (antibody or receptor) for ER retention, which would then limit the modification of the glycoprotein of interest.

Definitions

Dicotyledon (dicot): a flowering plant whose embryos have two seed halves or cotyledons. Examples of dicots include: tobacco; tomatoes; potatoes, the legumes including alfalfa and soybeans; oaks; maples; roses; mints; squashes; daisies; walnuts; cacti; violets; and buttercups.

Monocotyledon (monocot): a flowering plant whose embryos have one cotyledon or seed leaf. Examples of monocots include: lilies; grasses; corn; rice, grains including oats, wheat and barley; orchids; irises; onions and palms.

Lower plant: any non-flowering plant including ferns, gymnosperms, conifers, horsetails, club mosses, liver warts, hornworts, mosses, red algae, brown algae, gametophytes, sporophytes of pteridophytes, and green algae.

Structural gene: a gene coding for a polypeptide and being equipped with a suitable promoter, termination sequence and optionally other regulatory DNA sequences, and having a correct reading frame.

Signal sequence: a DNA sequence coding for an amino acid sequence attached to the polypeptide which binds the polypeptide to the endoplasmic reticulum and is essential for protein secretion.

Promoter: a recognition site on a DNA sequence or group of DNA sequences that provide an expression control element for a gene and to which RNA polymerase specifically binds and initiates RNA synthesis (transcription) of that gene.

Inducible promoter: a promoter where the rate of RNA polymerase binding and initiation is modulated by external stimuli. Such stimuli include light, heat, anaerobic stress, alteration in nutrient conditions, presence or absence of a metabolite, presence of a ligand, microbial attack, wounding and the like.

Viral promoter: a promoter with a DNA sequence substantially similar to the promoter found at the 5' end of a viral gene. A typical viral promoter is found at the 5' end of the gene coding for the p2I protein of MMTV described by Huang et al., 27(2) CELL 245-55 (1981).

Synthetic promoter: a promoter that was chemically synthesized rather than biologically derived. Usually synthetic promoters incorporate sequence changes that optimize the efficiency of RNA polymerase initiation.

Constitutive promoter: a promoter where the rate of RNA polymerase binding and initiation is approximately constant and relatively independent of external stimuli.

Temporally regulated promoter: a promoter where the rate of RNA polymerase binding and initiation is modulated at a specific time during development. Examples of temporally regulated promoters are given in Chua et al., 244 SCIENCE 174-81 (1989).

Spatially regulated promoter: a promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism such as the leaf, stem or root. Examples of spatially regulated promoters are given in Chua et al., 244 SCIENCE 174-81 (1989).

Spatiotemporally regulated promoter: a promoter where the rate of RNA polymerase binding and initiation is modulated in a specific structure of the organism at a specific time

during development. A typical spatiotemporally regulated promoter is the EPSP synthase-35S promoter described by Chua et al., 244 SCIENCE 174-81 (1989).

Heterologous Polypeptide: a linear series of amino acid residues connected one to the other by peptide bonds between the alpha-amino and carboxy groups of adjacent residues originating from a species other than the plant host system within which said linear series is produced. "Polypeptide" also encompasses a sequence of amino acids, peptides, fragments of polypeptides, proteins, globular proteins, glycoproteins, post-translational enzymes and fragments of these.

Multimeric protein: a protein containing more than one separate polypeptide or protein chain, each associated with the other to form a single protein. Both heterodimeric and homodimeric proteins are multimeric proteins.

Single-chain antigen-binding protein: a polypeptide composed of an immunoglobulin light-chain variable region amino acid sequence (V_L) tethered to an immunoglobulin heavy-chain variable region amino acid sequence (V_H) by a peptide that links the carboxyl terminus of the V_L sequence to the amino terminus of the V_H sequence.

Single-chain antigen-binding protein-coding gene: a recombinant gene coding for a single-chain antigen-binding protein.

Immunoglobulin: a polypeptide, protein or multimeric protein containing at least the immunologically active portion of an immunoglobulin heavy chain and is thus capable of specifically combining with an antigen. Exemplary immunoglobulins are immunoglobulin heavy chains, immunoglobulin molecules, substantially intact immunoglobulin molecules, any portion of an immunoglobulin that contains the paratope, including those portions known in the art as Fab fragments, Fab' fragment, F(ab').sub.2 fragment and Fv fragment.

Immunoglobulin molecule: a multimeric protein containing the immunologically active portions of an immunoglobulin heavy chain and immunoglobulin light chain covalently coupled together and capable of specifically combining with antigen.

Fab fragment: a multimeric protein consisting of the portion of an immunoglobulin molecule containing the immunologically active portions of an immunoglobulin heavy chain and an immunoglobulin light chain covalently coupled together and capable of specifically combining with antigen. Fab fragments are typically prepared by proteolytic digestion of substantially intact immunoglobulin molecules with papain using methods that are well known in the art. However, a Fab fragment may also be prepared by expressing in a suitable host cell the desired portions of immunoglobulin heavy chain and immunoglobulin light chain using methods well known in the art.

F_v fragment: a multimeric protein consisting of the immunologically active portions of an immunoglobulin heavy chain variable region and an immunoglobulin light chain variable region covalently coupled together and capable of specifically combining with antigen. F_v fragments are typically prepared by expressing in suitable host cell the desired portions of immunoglobulin heavy chain variable region and immunoglobulin light chain variable region using methods well known in the art.

Asexual propagation: Producing progeny by regenerating an entire plant from leaf cuttings, stem cuttings, root cuttings, single plant cells (protoplasts) and callus.

Glycosylated core portion: The pentasaccharide core common to all asparagine-linked oligosaccharides. The pentasaccharide core has the structure Manα1-3(manα1-6) Manβ1-46LcNAcβ1-4 6LcNac-(ASN amino acid). The pentasaccharide core typically has 2 outer branches linked to the pentasaccharide core.

N-acetylglucosamine containing outer branches: The additional oligosaccharides that are linked to the pentasaccharide core (glycosylated core portion) of asparagine-linked oligosaccharides. The outer branches found on both mammalian and plant glycopolypeptides contain N-acetylglucosamine in direct contrast with yeast outer branches that only contain mannose. Mammalian outer branches have sialic acid residues linked directly to the terminal portion of the outer branch.

Glycopolypeptide multimer: A protein containing a glycosylated polypeptide or protein chain and at least one other polypeptide or protein chain bonded to each other to form a single globular protein. Both heterodimeric and homodimeric glycoproteins are multimeric proteins. Glycosylated polypeptides and proteins are n-glycans in which the C(I) of N-acetylglucosamine is linked to the amide group of asparagine.

Immunoglobulin superfamily molecule: a molecule that has a domain size and amino acid residue sequence that is significantly similar to immunoglobulin or immunoglobulin related domains. The significance of similarity is determined statistically using a computer program such as the Align program described by Dayhoff et al., 91 METH. ENZYMOL. 524-45 (1983). A typical Align score of less than 3 indicates that the molecule being tested is a member of the immunoglobulin gene superfamily. The immunoglobulin gene superfamily contains several major classes of molecules described by Williams and Barclay, in IMMUNOGLOBULIN GENES 361 (Academic Press, New York, N.Y. 1989).

Plant culture medium: any combination of amino acids, salts, sugars, plant growth regulators, vitamins, and/or elements and compounds that will maintain and/or support the growth of any plant, plant cell, or plant tissue. A typical plant culture medium has been described by Murashige & Skoog, 15 PHYSIOL. PLANT. 473-97 (1962).

Plant transformation and cell culture: broadly refers to the process by which plant cells are genetically altered and transferred to an appropriate plant culture medium for maintenance, further growth, and/or further development.

Isolated: as used herein, refers to any element or compound separated not only from other elements or compounds that are present in the natural source of the element or compound, but also from other elements or compounds and, as used herein, preferably refers to an element or compound found in the presence of (if anything) only a solvent, buffer, ion, or other component normally present in a solution of the same.

Operably linked: as used herein, refers to the state of any compound, including but not limited to deoxyribonucleic acid, when such compound is functionally linked to any promoter.

Recombinant: as used herein, broadly describes various technologies whereby genes can be cloned, DNA can be sequenced, and protein products can be produced. As used herein, the term also describes proteins that have been produced following the transfer of genes into the cells of plant host systems.

Amino acid sequences: as used herein, this term includes an oligopeptide, peptide, polypeptide, or protein sequence, and fragment thereof, and to naturally occurring or synthetic molecules.

Collagen: includes any one of the collagen types, including collagen types I through XX, as well as any novel collagens whether natural, synthetic, semi-synthetic or recombinant. The term also encompasses both procollagen and mature collagen assembled as hetero- and homotrimers, and at least one polypeptide chain or subunit or fragment of collagen for any of the collagen types, and any heterotrimers of any combination of the collagen constructs of the invention. The term "collagen" is further defined to include other forms of the polypeptide that may be categorized as a subset to collagen, such as gelatins. The term "collagen" is meant to encompass all of the foregoing, unless the context dictates otherwise.

Collagen Type I: a major fibrillar collagen of bone and skin; a heterotrimeric molecule comprising two α1(I) chains and one α2(I) chain, encoded by the COL1A1 and COL1a2 genes, respectively.

Collagen Type II: the predominant collagen in cartilage and vitreous humor; a homotrimeric molecule comprising three identical α1(II) chains encoded by the COL2A1 gene.

Collagen Type III: a major fibrillar collagen found in skin and vascular tissues; a homotrimeric collagen comprising three identical α1(III) chains encoded by the COL3A1 gene.

Collagen Type IV: a collagen found in basement membranes in sheet form; the most common form of this collagen contains two 1(IV) chains and one 2(IV) chain.

Collagen Type V: a fibrillar collagen found primarily in bones, tendon, cornea, skin, and blood vessels; exists in both homotrimeric and heterotrimeric forms; may consist of two 1(V) chains and α2(V); may consist of α1(V), α2(V), and α3(V); may be a homotrimer of α1(V).

Collagen Type VI: a collagen found in connective tissue; a heterotrimer comprising α1(VI), α2(VI), and α3(VI) chains.

Collagen Type VII: a collagen found in particular epithelial tissues; a homotrimeric molecule of three α1(VII) chains.

Collagen Type VIII: a collagen found in Descemet's membrane in the cornea; a heterotrimeric molecule comprising two α1(VIII) chains and one α2(VIII) chain; a heterotrimer comprising other configurations reported and unreported.

Collagen Type IX: a fibril-associated collagen found in cartilage and vitreous humor; a heterotrimeric molecule consisting of α1(IX), α2(IX), and α3(IX) chains.

Collagen Type X: a collagen isolated from, among other tissues, hypertrophic cartilage found in growth plates; a homotrimer comprising α1(X) chains.

Collagen Type XI: a collagen found in cartilaginous tissues; a collagen associated with Collagen Type II and Collagen Type IX; a heterotrimer comprising α1(XI), α2(XI), and α(XI) chains.

Collagen Type XII: a fibril-associated collagen closely associated with Collagen Type I; a homotrimeric molecule comprising three α1(XII) chains.

Collagen Type XIII: a non-fibrillar collagen found, among other places, in skin, intestine, bone, cartilage, and striated muscle; detailed descriptions of this collagen appear in Juvonen et al., 267(34) J. BIOL. CHEM. 24700-07 (1992).

Collagen Type XIV: a fibril-associated collagen; a homotrimeric molecule comprising three α1(XIV) chains.

Collagen Type XV: a collagen homologous in structure to Collagen Type XVIII; information about the structure of Collagen Type XV is available in Myers et al., 89(21) P.N.A.S. 10144-48 (1992); Huebner et al., 14(2) GENOMICS 220-24 (1992); Kivirikko et al., 269(7) J. BIOL. CHEM. 4773-79 (1994); Muragaki et al., 269(6) J. BIOL. CHEM. 4042-46 (1994).

Collagen Type XVI: a fibril-associated collagen found in skin, lung fibroblasts, keratinocytes, and elsewhere; information about the structure of Collagen Type XVI is available in Pan et al., 89(14) P.N.A.S. 6565-69 (1992); Yamaguchi et al., 112(6) J. BIOCHEM. 856-63 (1992).

Collagen Type XVII: a hemidesmonal transmembrance collagen; information about its structure appears in Li et al., 268(12) J. BIOL. CHEM. 8825-34 (1993); McGrath et al., 11(1) NAT. GENET. 83-86 (1995).

Collagen Type XVIII: a collagen found in the liver; information about its structure appears in Rehn et al., 91(10) P.N.A.S. 4234-38 (1994); Oh et al., 91(10) P.N.A.S. 4229-33 (1994); Rehn et al., 269(19) J. BIOL. CHEM. 13929-35 (1994); Oh et al., 19(3) GENOMICS 494-49 (1994).

Collagen Type XIX: a fibril-associated collagen; information about its structure appears in Inoguchi et al., 117(1) J. BIOCHEM. (TOKYO) 137-46 (1995); Yoshioka et al., 13(3) GENOMICS 884-86 (1992); Myers et al., 269(28) J. BIOL. CHEM. 18549-57 (1994).

Collagen Type XX: a fibril-associated collagen found in chick cornea.

Collagen subunit: as used herein, the amino acid sequence of one polypeptide chain of a collagen protein encoded by a single gene, as well as derivatives, including deletion derivatives, conservative substitutions, etc.

Fusion protein: a protein in which peptide sequences from different proteins are covalently linked together.

Post-translational modification: includes modifications which occur after a nucleic acid has been translated into a sequence of amino acids. These modifications may either add and/or remove moieties from specific sites on the amino acid sequence. Example of specific sites present on a sequence of amino acids, include, but are not limited to, proteolytic cleavage sites, glycosylation sites, carbohydrate cleavage or substitution sites, phosphorylation sites, methylation sites, sulfation sites, prenylation sites, acetylation sites, N-amidation sites, and myristylation sites.

Post-translational modifying enzymes: enzymes responsible for either adding and/or deleting moieties from specific sites on the amino acid sequence. Examples of post-translation modifying enzymes include, but are not limited to glycoprotein glycosyltransferases, GlcNAc-1-phosphotransferase, GlcNAc 1-phosphodiester-N-acetylglucosaminindase, glycosidases, oxylases, exoglycosidases, endoglycosidases, GlcNAc phosphotransferase, protein kinases, 3'-phosphoadenosyl-5'-phosphosulphatase, prolyl hydroxylase, and lysyl hydroxylase. This definition includes any fragment, functional equivalent, or catalytic site, with equivalent or substantially equivalent activity of the native or authentic enzyme.

Glycosyltransferases: include a wide variety of post-translational enzymes that process peptides into glycoproteins. These include such enzymes as N-acetylglucosaminal transferases, N-acetylgalactosaminyltransferases, sialyl-transferases, fucosyltransferases, galactosyltransferases and mannosyltransferases. Other families of enzymes including sulfotransferases, glycosidases, acetyltransferases, and mannosidases also play various roles in processing particular glycoproteins. *See* U.S. Pat. No. 5,032,519.

Collagen post-translational enzyme: any enzyme that modifies a procollagen, collagen, or components comprising a collagen molecule, including but not limited to prolyl 4-hydroxylase, C-proteinase, N-proteinase, lysyl hydroxylase, and lysyl oxidase.

Hybridization: broadly defined, any process by which a nucleic acid sequence binds to a complementary sequence through base pairing. Hybridization conditions can be defined by, for example, the concentrations of salt or formamide in the prehybridization and hybridization solutions, or by the hybridization temperature, and are well known in the art. Hybridization can occur under conditions of various stringency. In particular, stringency can be increased by reducing the concentration of salt, increasing the concentration of formamide, or raising the hybridization temperature. For example, hybridization under high stringency conditions could occur in about 50% formamide at about 37°C to 42°C. Hybridization could occur under reduced stringency conditions in about 35% to 25% formamide at about 30°C to 35°C. In particular, hybridization could occur under high stringency conditions at 42°C in 50% formamide, 5 times SSPE, 0.3% SDS, and 200 ug/ml sheared and denatured salmon sperm DNA. Hybridization could occur under reduced stringency conditions as described above, but in 35% formamide at a reduced temperature of 35°C. The temperature range corresponding to a particular level of stringency can be further narrowed by calculating the purine to pyrimidine ratio of the nucleic acid of interest and adjusting the temperature accordingly. To remove nonspecific signals, blots can be sequentially washed, for example, at room temperature under increasingly stringent conditions of up to 0. 1 X saline sodium citrate and 0.5% sodium dodecyl sulfate. Variations on the above ranges and conditions are well known in the art.

Altered nucleic acid sequences: as the term is used herein, nucleic acid sequences encoding a heterologous polypeptide, post-translational modifying enzyme, or functional equivalent thereof including those sequences with deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent post-translational enzyme or heterologous polypeptide. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding a post-translational enzyme or heterologous polypeptide and improper or unexpected hybridization to alleles, with a locus

other than the normal chromosomal locus for the polynucleotide sequence encoding a post-translational enzyme or heterologous polypeptide. The encoded protein may also be "altered" and contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent post-translational enzyme or heterologous polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological or immunological activity of a post-translational enzyme or heterologous peptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine, glycine and alanine, asparagine and glutamine, serine and threonine, and phenylalanine and tyrosine.

Antisense gene: an antisense gene is constructed by reversing the orientation of the gene with respect to its promoter so that the antisense strand is transcribed.

Antisense RNA: an RNA molecule complementary to a particular RNA transcript that can hybridize to the transcript and block its function.

Fragments: include any portion of an amino acid sequence which retains at least one structural or functional characteristic of the subject post-translational enzyme or heterologous polypeptide.

Introduction: insertion of a nucleic acid sequence into a cell, by methods including infection, transfection, transformation or transduction.

Functional equivalent: a protein or nucleic acid molecule that possesses functional or structural characteristics that are substantially similar to a heterologous protein, polypeptide, enzyme, or nucleic acid. A functional equivalent of a protein may contain modifications depending on the necessity of such modifications for the performance of a specific function. The term "functional equivalent" is intended to include the "fragments," "mutants," "hybrids," "variants," "analogs," or "chemical derivatives" of a molecule.

Chemical derivative: as used herein, a molecule is said to be a "chemical derivative" of another molecule when it contains additional chemical moieties not normally a part of the molecule. Such moieties can improve the molecule's solubility, absorption, biological half-life, and the like. The moieties can alternatively decrease the toxicity of the molecule, eliminate or attenuate any undesirable side effect of the molecule, and the like.

Variant: an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Analogous minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted may be found using computer programs well known in the art, for example, DNASTAR© software.

Vector: a cloning vector that is designed so that a coding nucleic acid sequence inserted at a particular site will be transcribed and translated. A typical expression vector may contain a promoter, selection marker, nucleic acids encoding signal sequences, and regulatory sequences, e.g., polyadenylation sites, 5'-untranslated regions, and 3'-untranslated regions, termination sites, and enhancers. "Vectors" include viral derived vectors, bacterial derived vectors, plant derived vectors and insect derived vectors.

Transformation: a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the type of host cell being transformed and may include, but is not limited to, viral infection, electroporation, heat shock, lipofection, and particle bombardment.

Protein purification: broadly defined, any process by which proteins are separated from other elements or compounds on the basis of charge, molecular size, or binding affinity.

Transgenic plants: plant host systems that have been subjected to one or more methods of genetic transformation; plants that have been produced following the transfer of genes into the cells of plant host systems.

Glycosylation

Glycosylation is a form of post-translational modification by which sugar residue chains are added onto peptides as they pass through the endoplasmic reticulum (ER) and Golgi subcellular compartments. The sugar, or glycan, chains of these glycoproteins are biologically important in at least two ways. For example, the sugar chains may function as ligands for binding cells and as receptors for bacteria and viruses. Indeed, the HIV envelope glycoprotein gp120 targets receptors on CD4 cells. Note that the normal function of such a receptor is to recognize the sugar residues of foreign glycoproteins to effect clearance from the body. Rahebi et al., 12 GLYCOCON. J. 7-16 (1995). Another function of the sugar chain is in the formation of higher order proteins, dissolution, protease resistance, antigenic inhibition, protein function modification, protein regeneration rate adjustment, and adjustment of the amount of glycoprotein expressed in cell layers. For example, sugar chains are instrumental in the adjustment of the adhesion of nerve cell molecules which are distributed widely in the nervous system. *See* Edelman, 54 ANN. REV. BIOCHEM. 135-69 (1985). Several of these aspects are discussed in more detail below.

Plant-produced glycoproteins have glycan patterns that often vary from their human counterparts. In some cases, removal of the glycan portion of the protein may be advantageous. For example, the first antibody produced in plants for injection into humans, NeoRx Corporation's PRETARGET® antibody (NeoRx Corporation, Seattle, WA) is non-glycosylated. However, occasions may arise in which a glycosylated antibody is desired. The same is true for other pharmaceutically important glycosylated proteins. Although plants save significant capital, costs, and yield processing advantages over mammalian systems in providing glycosylated proteins, the differences between plant and mammalian glycosylation may create a limit for some plant applications. The present invention overcomes such limitations by providing plants that produce mammalian-like glycosylation to heterologous proteins produced in plants.

By way of background, the asparagine-linked glycan portion of glycoproteins share a common core structure, for which the basic biochemical process is conserved, in plants, mammals and other higher organisms (eukaryotes). Specifically, asparagine-linked glycosylation, initiated in all eukaryotes at the conserved signal (Asn-Xxx-Thr/Ser), starts with the addition of an identical glycan structure in all higher organisms. Subsequent glycosylation steps can modulate both the glycoprotein structure and the functionality it imparts to the protein.

Generally, N-glycosylation of proteins in plant cells occurs via the following pathway. An oligosaccharide precursor, GlcManGlcNAc, is transferred from a oligosaccharyl-pyrophosphoryl-dolichol carrier to an asparagine on the nascent protein. The asparagine reside is generally part of the peptide sequence: Asn-X-Ser/Thr, where X is any residue except proline, asparatate, or glutamate. The terminal glucose units of the oligosaccharide are then removed by glucosidase I and II. In mammals, a mannose residue is also removed from the oligosaccharide by a mannosidase located in the ER. However, an ER-specific mannosidase has not been detected in plants. Rayon et al., 49 J. EXP. BOTANY 1463-72 (1998).

The N-glycans of proteins may be further modified in the Golgi during transport from the cis to the trans cisternae. Initially, a N-acetylglucosamine residue is added to the mannose residue of the N-glycan catalyzed by N-acetylglucosaminyltransferase I (GNT I). Johnson et al., 84 PLANT PHYSIOL. 1301-08 (1987). Following this addition, two additional mannose residues are removed by mannosidase II and an N-acetylglucosamine residue is transferred by GNT II. Tezuka et al., 203 EUR. J. BIOCHEM. 401-13 (1992). The N-glycan may also undergo fucosylation and xylosylation. Furthermore, the N-glycan may be further modified with the addition of terminal fucose and galactose residues to produce mono- and biantennary plant complex N-glycans. Fitchette-Lainé et al., 12 PLANT J. 1411-17 (1997).

These subsequent steps in the glycosylation pathway are not always similar in all hosts, yielding a difference between the standard human glycan pattern and that from the production host source (e.g., mammalian cell, cow milk, yeast, or plant). *See* Jenkins et al., 14 NATURE BIOTECH. 975-81 (1996); Goochee et al., 9 BIOTECH. 1347-55 (1991); Lerouge et al., 38 PLANT MOL. BIOL. 31-48 (1998); and U.S. Pat. No. 5,202,422. However, the modification of the core structure by the addition and removal of sugar units can be very different in different host species. A subset of the possible structures for mammals is shown

in Fig. 1, and for plants in Fig. 2. These glycan patterns are created by the addition and subtraction of different sugars by specific enzymes present in the different systems. The resultant glycan structure can influence some of the activities of the therapeutic glycoprotein. Regarding antibodies, *see* Wright & Morrison, 15 TIBTECH 26-32 (1997).

Many immunoglobulin gamma class antibodies (IgG), require particular glycosylation at the conserved CH2 domain of the heavy chain for complete functionality, such as antibody dependent cell-mediated cytotoxicity (ADCC) and complement-mediated cytotoxicity (CMC). The presence of glycosylation at the antigen binding domain, although rare, may sometimes influence antigen binding. Glycosylation presence or structure may also influence protease sensitivity, receptor binding, and serum half-life of some glycoproteins. Moreover, particular types of sugar residues, such as neuraminic acid, galactose, N-acetylglucosamine, mannose, fucose and xylose, play different roles in protein character.

For example, although some antibodies from mammalian systems contain neuraminic acid (sialic acid) termini, this is apparently not required on IgG for full functionality. *See* Sheeley et al., 247 ANAL. BIOCHEM. 102-10 (1997); Parekh et al., 316 NATURE 452-57 (1985); Weitzhandler et al., 83 J. PHARMA. SCI. 1670-75 (1994). For some proteins, such as erythropoeitin, neuraminic acid addition at the ends of the glycan chains is important to ensure half-life in sera, but removal may increase *in vitro* activity. Goochee et al., 1991. Plants have not previously been reported to either make neuraminic acid, or add this sugar to the glycan core.

Regarding another sugar residue, galactose, *in vitro* and *in vivo* testing has shown that galactose ends are critical to recruit ADCC and CMC functions to the site of antibody-antigen interaction. These functions are important for some, but not all, therapeutic antibodies. For some glycoproteins, galactose termini may signal turnover via the galactose receptor. *See* 333-35 Geisow, 10 TIBTECH (1992). This may not be the case for antibodies, as sera contains high levels of antibodies with these ends. Plants have the capacity to make galactose ends, but there is conflicting data concerning how commonly it occurs, and at which linkage. Terminal galactose as shown in Fig. 2 (c and d) has not been observed previously in antibodies produced from plants.

Additionally, some evidence indicates that IgG with glycan chains ending in N-acetyl glucosamine (GlcNac) not only lose ADCC and CMC functionality, but such ends may also correlate with rheumatoid arthritis. This has been surmised from both population studies

(Parekh et al., 1985) and biochemical studies (Malhotra et al., 3 NATURE MED. 237-43 (1995), and is thought to be due to capture of the antibody in joint areas by a mannose-type receptor. Reports show a percentage of plant glycoproteins end in GlcNac, but this has not been observed previously in glycosylated antibodies produced in plant systems.

Further truncation of antibodies to mannose ends has been done in non-plant systems by genetic (Wright & Morrison, 180 J. EXP. MED. 1087-96 (1994)) and enzymatic (Reinholdt et al., 58 INFECTION & IMMUNITY 1186-94 (1990)) means. In the former case, the results indicate a lower half-life that may be caused by antibody binding to a mannose-type receptor. No ADCC or CMC activity was observed, and the effect on rheumatoid arthritis was not directly tested. Mannose termini are often seen in plant glycoproteins. Additionally, fucose addition to a subset of a glycoprotein population has been noted in all mammalian systems, but does not appear to have a function. The particular fucose linkage on such glycoproteins differs between plants and animals. The terminal fucose shown in Fig. 2 (c and d) has not been observed previously in antibodies produced in plants.

Production of glycosylated proteins in plants is advantageous over mammalian systems that also produce glycoproteins that vary from their human counterparts. In general, therapeutic glycoproteins from mammalian cell production systems can have a similar range of glycan forms to that seen on the comparable protein from humans, but the molar ratios of the glycan forms are different. There can also be notable non-human glycan forms from cell systems, such as structures d and e in Fig. 1. *See* Jenkins et al., 1996. For example, erythropoietin from both human urine and CHO cells both have significant levels of tetraantennary glycans, but the CHO source has a much higher fraction with an additional N-acetyllactosamine repeat, and does not add any neuraminic acid in an α2-6 linkage. *See* Takeuchi et al., 263 J. BIOL. CHEM. 3657-63 (1988). The higher incidence of novel neuraminic acid forms on four different recombinant proteins from CHO cells was also noted by Hokke et al., 1990, 275 FEBS LETTERS 9-14 (1990). Finally, comparisons between mouse milk, CHO cell, and insect cell interferon production indicate differences in how often one of two glycosylation sites are used, and what type of glycans are present. *See* James et al., 13 BIOTECH. 592-96 (1995). Hence, using plants to manufacture these glycoproteins such that more human-like glycosylations are achieved is an advantage of the present invention over other strategies.

In the case of antibodies, these proteins tend to have simpler, biantennary glycans, the most highly studied being in the conserved CH2 domain of the IgG antibody class. Other glycosylation sites, such as in the variable region (binding the antigen) are seen less frequently. *See* Co et al., 30 MOLEC. IMMUN. 1361-67 (1993); Tachibana et al., 16 CYTOTECH. 151-7 (1994); Wallick et al., 168 J. EXP. MED. 1099-109 (1988). The glycan structures on antibodies collected from pooled human sera and from non-human mammalian systems (monoclonal antibodies) have been compared. Weitzhandler et al., (1994) indicated a human IgG pool has a higher fraction with galactose ends than either a mouse cell- or mouse ascites-derived monoclonal. A novel form of neuraminic acid was also noted, but neuraminic acid is not a common end of IgG glycan chains from any sources tested. Sheeley et al., (1997) noted that CHO cell production of a monoclonal more approximated the glycosylation pattern seen on pooled human sera, when compared to NSO cells. Most notably, the NSO cells added some level of an extra terminal galactose, with a novel α1-3 linkage. This was also noted by Lund et al., 27 MOL. IMMUNOLOGY 1145-53 (1990), in a different mouse cell system. Humans have high levels of antibodies targeting this foreign linkage, which may limit therapeutic efficacy. *See* Borrebaek et al., 14 IMMUNOLOGY TODAY 477-79 (1993). These authors noted the novel galactose ends may limit serum half-life, but did not mention the potential for anaphylaxis. Thus, the ability of plants to produce antibodies with more human-like glycosylations is desirable to avoid this reactivity.

Moreover, plant production of glycoproteins may overcome the disadvantage whereby the culture conditions of mammalian cell systems influence the glycan structure of the resulting glycoproteins. For example, low glucose culture concentrations can increase levels of alternately glycosylated light chain, and change antibody binding characteristics (Tachibana et al., 1994). Heavy chain behavior was not analyzed by this group. Similarly, during a CHO cell culturing period the proportion of interferon with neither glycan site filled rises steadily from 0% to 30%, depending on external glucose levels tested. Curling et al., 272 BIOCHEM. J. 333-37 (1990). Partial glycosylation is also seen when prolactin is produced from murine C127 cells; although both forms are also seen in normal human sera. Price et al., 136 ENDOCRINOLOGY 4827-33 (1995). Other variables are reviewed by Goochee & Monica, 8 BIOTECH. 421-26 (1990) and Jenkins et al., (1996), including amine concentrations, hormones, pH, buffering components, specific drug inhibitors, time, and other cell culture conditions. Some of these factors may be relevant in production settings; a

mechanism for each is not always known. Hence, overcoming the variability present in mammalian culture by providing plant systems that produce uniform populations of glycoproteins is another advantage of the present invention.

As noted, previous recombinant plant-source glycoproteins have, like mammalian systems, exhibited glycans different from human-source glycoproteins. Compared to mammalian systems, however, there are fewer detailed studies of recombinant therapeutic glycoproteins from plants. Extrapolations from the few recombinant studies, and many endogenous (non-recombinant) studies, indicate a core structure common with mammals, with some novel glycan additions, some range in frequency of glycosylation site utilization, and a narrower range in types of structures. This was reviewed by Lerouge et al., 38 PLANT MOLEC. BIOL. 31-48 (1998).

Further comparing animal versus plant systems, following initial glycan addition, the subsequent residue trimming and GlcNac addition is as in mammals, but the GlcNac is removed at a higher frequency in plants. *See, e.g.*, Sturm, 199 EUR. J. BIOCHEM. 169-79 (1991); Hayashi et al., 191 EUR. J. BIOCHEM. 287-95 (1990); Sturm et al., 262 J. BIOL. CHEM. 13392-403 (1987). In plants, significant levels of glycoproteins can sometimes be found that have not progressed through the later stages of glycan maturation, yielding high mannose structures. Hayashi et al., 1990; Lerouge et al., 38 PLANT MOLEC. BIOL. 31-48 (1998). The next sugar to be added in a mammalian system, galactose, is rarely seen in plants. Faye et al., 75 PHYSIOL. PLANTARUM 309-14 (1989); Lerouge et al., 1998. This sugar may be important for full antibody complement activity. Finally, plants do not seem to make or add any neuraminic acid (Faye et al., 1989), a sugar not apparently critical for antibody function. However, this deficiency would help explain the behavior of erythropoietin from tobacco cells: novel size, poor secretion, expected *in vitro* activity, but no *in vivo* activity. Matsumoto et al., 27 PLANT MOLEC. BIOL. 1163-72 (1995). Published data with a recombinant antibody from tobacco also suggests a lack of both neuraminic acid and galactose. Hein et al., 7 BIOTECHNOL. PROG. 455-61 (1991). Data from tobacco cells, corn, and soy are consistent with this view, as well as greatly reduced levels of GlcNac on the glycan branches. This sort of mannose-terminating structure is not noted in mammalian sources of antibodies (Sheeley et al., 1997; Parekh et al., 1985; Weitzhandler et al., 1994), although a low level of mannose ends is seen in CHO-produced interferon (James et al., 1995).

As with other host systems (e.g., mouse, insect, yeast), plant systems also have the capacity to add novel sugars. In all plant systems studied, a novel fucose linkage and/or the presence of xylose has been noted (Sturm, 1991; Hayashi et al., 1990; Sturm et al., 1987). Limited work shows the same is true for recombinant proteins from plants. These novel sugars have been shown to be antigenic when natural plant proteins are used as immunogens in rabbits (Faye et al., 1989). More importantly, testing of sera from plant allergy patients indicate a positive reaction to these novel plant sugars. Garcia-Cassado et al., 6 GLYCOBIOLOGY 471-77 (1996). In contrast, initial work using recombinant humanized antibody as an immunogen in mice indicated a reaction only to the non-mouse peptide portion, and not the plant glycan. Minimally, the published results indicate that some patients with sensitivities to plants might be excluded from therapies with plant-derived injectable glycoproteins, until the advent of the invention herein, which provides for human-like glycosylations on plant-produced proteins.

The invention herein also addresses the consistency of production of therapeutic glycoproteins from plants by decreasing the variability of antibody structure. Although glycosylation versus growth and environment has not been as rigorously tested in plant systems as in mammalian systems, some available data show which factors have influence. All published studies involve endogenous proteins. Rice seedlings yield a different distribution of glycans on a model protein than seen in rice cell culture; the culture was also influenced by the presence of primary amines, as in mammalian cultures. Hayashi et al., 1990. The glycan structure at one site in a bean protein is influenced by the presence or absence of a glycan at a second site. Strum et al., 1987. As with different mammalian systems, the same protein put into different plant systems can yield different ratios of glycan forms. Lerouge et al., 1998. This pattern can be simplified in plants (von Schaewen et al., 102 PLANT PHYSIOL. 1109-118 (1993)) by a mutation in the same enzyme shown to yield high mannose ends in mammalian systems. Wright & Morrison, 1994. This mutation eliminates the variability, as well as the novel fucose and xylose residues, seen in plants.

Until the advent of the invention herein, and given the variability of outcome regarding non-human-like glycosylation in clinical outcomes, one alternative to generate less variable glycoproteins may be to eliminate the glycosylation. This would be most practicable with antibodies, typically having only one site of glycosylation. As expected from the above discussion, any antibody de-glycosylation method tested reduced the level of *in vitro* effector

function activity. *See* Pound et al., 30 MOL. IMMUNOLOGY 233-41 (1993); Tao & Morrison, 143 J. IMMUNOLOGY 2595-601 (1989); Leader et al., 72 IMMUNOLOGY 481-85 (1991); Lund et al., 1990; Hand et al., 1992; Winkelhake et al., 255 J. BIOL. CHEM. 2822-28 (1980); Nose & Wigzell, 80 P.N.A.S. 6632-36 (1983); U.S. Pat. No. 5,648,260. Genetic, chemical, and enzymatic methods have been described to remove the glycan. Simple genetic mutations of a single amino acid are the most direct method; typically, the consensus site asparagine is changed to the related aspartate or glutamate. In the extreme case, Centocor (Malvern, PA) produces the REOPRO® anticlotting agent as a proteolyzed Fab antibody fragment, devoid of glycosylation and the terminal part of the heavy chain. The production of a genetically non-glycosylated antibody by Monsanto Company (St. Louis, MO) to support a clinical study by NeoRx indicated therapeutic equivalence to the mammalian-produced product having a modified glycan. Both showed similar tissue distribution and half-life in humans. An example of the similar clinical data of the mammalian and plant antibodies is shown in Fig. 3.

In addition to perturbations of antibody effector functions following changes in glycosylation, published reports show other influences of glycan removal may be more variable. For example, while *in vitro* protease sensitivity may be increased, animal half-life may remain fairly constant both in mice (Tao & Morrison, 143 J. IMMUNOL. 2595-601 (1989)), and primates. Hand et al., 35 CANCER IMMUNOL. & IMMUNOTHERAPY 165-74 (1992). Elimination of glycosylation at the antigen binding site can increase the affinity of the antibody for antigen. Co et al., 1993 and Wallick et al., 1988. Although there are concerns that glycan removal may cause reduced solubility (and subsequently perhaps increase antigenicity), these antibodies have been shown to be stable early in development and formulation, and have been soluble. Regarding other recombinant therapeutic proteins, granulocyte colony-stimulating factor (G-CSF), marketed as NEUPOGEN® by Amgen, Inc. (Thousand Oaks, CA), is refolded from bacteria, and so is devoid of the natural O-linked glycosylation. This same method of mutation was used to produce nonglycosylated urinary plasminogen activator from yeast. Melnick et al., 265 J. BIOL. CHEM. 801-07 (1990).

Developing host plants that do not add or maintain xylose and fucose on their glycoproteins may yield glycoproteins that are less antigenic. Developing host plants that maintain N-acetylglucosamine linkages, and are able to add glucose, may yield plant-produced antibodies that are more functional for some uses, such as complement-activating antibodies. Developing plants with more uniform glycosylation pattern may yield more

uniform therapeutic biologics produced in plants. Also, altering glycosylation may alter plant growth, and either reduce or increase susceptibility to stresses and pathogens as desired. The effect may be enhanced growth to improve yield or reduced growth to aid plant control, changed pollen interactions for genetic containment, or identification of novel herbicides related to the glycosylation pathway.

Post-Translational Modification of Collagen

Collagens are from a family of highly characterized fibrous proteins that are found in many multicellular animals. Collagen is one of the most abundant structural proteins in mammals, comprising about 25% of the total protein mass, and is present in skin, bones, cartilage, tendons, blood vessels, teeth, and most organs of the body, where it serves to hold cells together in discrete units.

Collagen in its native form is typically a rigid, rod-shaped molecule approximately 300 nm long and 1.5 nm in diameter. It is composed of three collagen polypeptides, α-chains, which are wound around each other to form a tight, ropelike, triple helix. The collagen polypeptides are characterized by a long midsection having the repeating sequence – Gly-X-Y, where X and Y are generally proline or hydroxyproline, bounded at each end by the telopeptide regions, which constitute less than about 5% of the collagen molecule. The telopetide regions of the collagen chains are typically responsible for the crosslinking that occurs between the chains, and for the immunogenicity of the protein. So far, about 25 distinct collagen α-chains have been identified, each encoded by a separate gene. Although in principle, combinations of those gene expressed could produce 10,000 types of triple-stranded collagen molecules, only about 20 types of collagen have been found. These collagen types are numbered by Roman numerals, and the chains found in each collagen type are identified with Arabic numerals. A detailed description of structure and biological functions of the various different types of naturally occurring collagens is provided in the art. *See* AYAD et al., THE EXTRACELLULAR MATRIX FACTS BOOK (Academic 1994); Burgeson & Nimmi, 282 CLIN. ORTHOP. 250-72 (1992); Kielty et al., in CONNECTIVE TISSUE AND ITS HERITABLE DISORDERS, MOLECULAR GENETICS, AND MEDICAL ASPECTS 103-47 (Royce & Steinmann eds., 1993).

Collagen has many useful properties, which are desirable in a number of medical, pharmaceutical, food, and cosmetic applications. For example, collagen is an important

component of bone grafts (Mulliken et al., PLASTIC AND RECONSTRUCTIVE SURGERY, May 1980, at 553; U.S. Pat. Nos. 6,077,988; 5,972368; 5,776,193; and 4,440,750); dermal implants (U.S. Pat. Nos. 4,448,911; 4,424,208; and 3,949,073); hemostats (Bell et al., 76(3) P.N.A.S. 1274-78 (1979); Weinberg et al., 122 J. CELL PHYSIOL. 410-14 (1985); Ehrlich, 82 J. CELL SCI. 281-94 (1986); Weinberg et al., 231 SCIENCE 397-400 (1986); U.S. Pat. Nos. 5,679,372; 4,891,359; 4,404,970; and 4,271,070); and incontinence implants (Shortliffe et al., 141 J. UROLOGY 538-541 (1989); U.S. Pat. Nos. 5,705,488; 5,523,291; 5,490,984; and 5,480,644). In addition, collagen is widely used in drug delivery systems. *See* Weiner et al., 74(9) J. PHARM. SCI. 922-25 (1985); U.S. Pat. Nos. 5,874,006; 5,807,591; and 4,789,662. Additionally, collagen is used in medical applications, such as the treatment for autoimmune disorders, *i.e.*, rheumatoid arthritis and has been evaluated in clinical trials for its potential for inducing oral-tolerance. *See* Trentham et al., 261 SCIENCE 1727-30 (1993); Thompson et al., 16 AUTOIMMUNITY 189-99 (1993); U.S. Pat. Nos. 6,083,918; 6,019,971; 6,010,722; 5,925,736; and 5,733,547. Collagen is also applied in food products such as sausage casings, which are derived from porcine, bovine, and sheep. *See* U.S. Pat. Nos. 5,840,849; 5,599,570. Finally, in health and beauty applications, collagen is frequently found in cosmetics or facial and skin products such as moisturizers. *See* U.S. Pat. Nos. 6,077,520; 6,036,966; and 4,942,153.

There are a variety of readily available procedures for preparing collagens from natural sources such as bovine hide, cartilage, and bones for these purposes. Generally, the bones are dried, defatted, crushed, and demineralized to extract collagen, while hides and cartilage are usually minced and digested with proteolytic enzymes. As collagen is resistant to most proteolytic enzymes, this procedure conveniently serves to remove most of the contaminating proteins found with collagen. Another procedure described by U.S. Pat. No. 3,949,173, involves the preparation of soluble collagen by dissolving animal tissue in aqueous acid, followed by enzymatic digestion. In addition, U.S. Patent No. 4,488,911 discloses a method for preparing collagen in solution (CIS), where the native collagen is extracted from animal tissue in dilute aqueous acid, followed by digestion with an enzyme, such as pepsin, trypsin, or Pronase®. The enzymatic digestion removes the telopeptide portions of the collagen molecules, providing atelopeptide collagen in solution. This atelopeptide is substantially non-crosslinked due to loss of the primary crosslinking regions.

Although collagen isolated with these procedures has been used with some success, there are some major disadvantages associated with these procedures. First, administering an animal-derived collagen to a human may elicit an undesired immunogenic response. Second, there are many safety issues that are of concern when using animal-derived collagen. The use of animal-derived collagens provide a potential risk for human exposure to pathogens and adventitious contaminants that are present in animal tissues. For example, several contaminants that are both difficult to detect and remove from animal tissues include scrapie, Transmissible Spongiform Encephalopathy agents (TSE), and Bovine Spongiform Encephalopathy (BSE). Rohwer, 88 DEV. BIOL. STAND. 247-56 (1996).

Gelatin, e.g., denatured collagen, is also a common and essential component for numerous applications in the pharmaceutical, medical, food, and photographic industries. For example, gelatin possesses many technological properties to support food preparations, and is used to thicken, emulsify, and improve consistency in foods such as jelly, breads, dairy, meat and fish products. Pharmaceutical applications of gelatin include hard and soft-shell capsules, microencapsulation, tabletting, suppositories, and medicinal emulsions. In addition, gelatin is an essential component of vaccine stabilizers. In the photography sector, gelatin is used to support silver halides on both film and paper. Varying characteristics and types of gelatin are desirable depending on the application and end product.

Indeed, from the disadvantages aforementioned, it is clear that it would be advantageous to supply human rather than animal-derived collagen and gelatin for medical, pharmaceutical, food, and cosmetic purposes. The sources of human collagen are limited and the only reliable sources, so far, are derived from human placenta and from the milk of transgenic animals. Human collagen can be purified from human placenta as described in U.S. Pat. No. 5,428,022. The placenta contains several types of collagen, most notably types I, III, IV, and V. Production of human placenta-derived collagen has several disadvantages associated with it as well. First, the process of separating and purifying one type from the others is imperfect and results in a predominant type with small amounts of the other types. Second, placenta-derived collagen requires additional processing steps to ensure that the resultant collagen product is free from human viruses, such as hepatitis and HIV.

Recombinant human collagen derived from the milk of transgenic animals is described in U.S. Pat. Nos. 5,667,839 and 6,111,165 and involves recovering human

procollagen or collagen from the mammary glands of a nonhuman mammal. The mammal will have been modified to contain an expression system that comprises DNA encoding procollagen under the control of regulatory sequences operable in mammary glands. This technique, however, has numerous shortcomings that include lengthy lead times to develop herds with stable genetics, high operating costs, inconsistent yields of recombinant protein, and contamination by animal viruses.

In view of these issues, it is desired to produce human collagen by using recombinant techniques that do not involve the use of transgenic animals. The production of recombinant human collagen, however, is troublesome due to the requirement of a plurality of post-translational modification enzymes, which are generally believed to be present only in cells which natively produce collagen. One of the essential post-translational steps is the conversion of specific proline residues to hydroxyproline, which ensures the stabilization of the triple helix at room temperature. Other post-translational modifications include disulfide exchange, hydroxylation of lysyl residues, addition of carbohydrates, and the assembly and crosslinking of the triple helical collagen molecules. Prockop et al., 311 NEW ENGL. J. MED. 276-89 (1984).

One of the shortcomings of using plant host systems to produce recombinant human collagen is that plants either do not naturally produce sufficient amounts of or do not have the post-translational enzymes needed to produce a biologically active recombinant protein, such as collagen. Therefore, one of the objectives of the present invention, is to supplement the plant host system with the post-translational enzymes required to produce biologically active recombinant heterologous proteins, such as collagen and therapeutic antibodies.

Post-translational enzymes are important to the biosynthesis of procollagens and collagens as with many other mammalian and human proteins. Collagen post-translational enzymes have been identified and reported in the literature, including prolyl 4-hydroxylase, C-proteinase, N-proteinase, lysyl oxidase, and lysyl hydroxylase. *See, e.g.,* OLSEN ET AL., CELL BIOLOGY OF EXTRACELLULAR MATRIX, (Hay ed., 2nd ed., Plenum Press, New York 1991).

More specifically, for example, prolyl 4-hydroxylase is a post-translational enzyme necessary for the synthesis of procollagen or collagen by cells. The enzyme is required to hydroxylate prolyl residues in the Y-position of the repeating -Gly-X-Y- sequences to 4-hydroxyproline. *See, e.g.,* Prockop et al., 311 NEW ENGL. J. MED. 376-86 (1994); U.S. Pat.

No. 5,593,859; WO 97/38710. Unless an appropriate number of Y-position prolyl residues are hydroxylated to 4-hydroxyproline by prolyl 4-hydroxylase, the newly synthesized procollagen chains cannot fold into a triple-helical conformation at 37°C. Moreover, if no hydroxylation or under-hydroxylation occurs, the polypeptides remain non-helical, are readily degraded, poorly secreted, and cannot self-assemble into collagen fibrils.

Prolyl 4-hydroxylase from vertebrates is an $\alpha_2\beta_2$ tetramer. *See, e.g.* Berg et al., 248 J. BIOL. CHEM. 1175-92 (1973); Tuderman et al., 52 EUR. J. BIOCHEM. 9-16 (1975). The α subunits (63 kDa) contain the catalytic sites involved in the hydroxylation of prolyl residues, but are insoluble in the absence of β subunits. The β subunits (55 kDa) were found to be identical to the protein disulfide isomerase. Protein disulfide isomerase catalyzes thiol/disulfide interchange in a protein substrate, leading to the formation of a set of disulfide bonds essential to establishing a stable protein. The β subunits retain 50 percent of protein disulfide isomerase activity when part of the prolyl 4-hydroxylase tetramer. *See, e.g.*, Pihlajaniemi et al., 6 EMBO. J. 643-49 (1987); Parkkonen et al., 256 BIOCHEM. J. 1005-11 (1988); Koivu et al., 262 J. BIOL. CHEM. 6447-49 (1987); U.S. Pat. No. 5,928,922. Active recombinant human prolyl 4-hydroxylase has been produced in insect cells by simultaneously expressing the α and β subunits in Sf9 cells. *See, e.g.*, Vuori et al., 89 P.N.A.S. 7467-70 (1992). An object of this invention is to provide a plant cell line that expresses post-translational enzymes able to effect the processing of collagen.

C-proteinase processes the assembled procollagen by cleaving off the C-terminal ends of the procollagens that assist in assembly of, but are not part of, the triple helix of the collagen molecule. *See generally*, Kadler et al., 262 J. BIOL. CHEM. 15969-71 (1987), Kadler et al., 580 ANN. NY ACAD. SCI. 214-24 (1990).

N-proteinase processes the assembled procollagen by cleaving off the N-terminal ends of the procollagens that assist in the assembly of, but are not part of, the collagen triple helix. *See generally*, Hojima et al., 269 J. BIOL. CHEM. 11381-90 (1994).

Post-Translational Modification of Human Protein C

Another aspect of the present invention is the production of fully functional Human Protein C in plants. Fully functional protein C is produced by several post-translational modifications. Two of the post-translational modification enzymes needed are the subtilisin-

like serine protease furin, also known as paired basic amino acid cleaving enzyme (PACE) and γ-glutamyl carboxylase.

Human protein C (HPC) is a complex plasma glycoprotein that functions as an anticoagulant by proteolytically inactivating coagulation factors V_a and $VIII_a$. The protein C anticoagulant pathway is triggered when thrombin binds to the endothelial cell receptor, thrombomodulin. This complex activates protein C to generate the anticoagulant enzyme, activated protein C (APC), which, in complex with protein S, inhibits coagulation by inactivating the critical regulatory proteins, factors V_a and $VIII_a$.

The protein C precursor is proteolytically modified to produce its mature form by the PACE. Specifically, endoproteolytic processing of the protein C precursor to generate its mature form involves cleavage of the propeptide after amino acids Lys^{-2}-$Arg4^{-1}$ and removal of a Lys^{156}-Arg^{157} dipeptide yielding the light and heavy chains of the mature protein C dimer.

Second, functional protein C is modified to allow binding to the endothelium cell protein C receptor (EPCR). Indeed, it is the vitamin K-dependent γ-carboxyglutamic acid (Gla) domain of activated protein C which allows the protein to bind to the EPCR. Specifically, the γ-carboxyglutamate residues of functional protein C are carboxylated by the enzyme, γ-glutamyl carboxylase, to form sites with high affinity for Ca^{2+}. The binding of Ca^{2+} by functional protein C allows it to readily bind to the EPCR.

Plant Expression Vectors

Expression vectors useful in the present invention comprise a chimeric gene (or expression cassette), designed for operation in plants, with companion sequences upstream and downstream from the expression cassette. The companion sequences may be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to be generated in bacteria and then introduced to the desired plant host.

The basic bacterial/plant vector construct will preferably comprise a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T-DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant

cell has been transformed. A general review of suitable markers for the members of the grass family is found in Wilmink & Dons, 11(2) Plant MOL. BIOL. REPTR. 165-85 (1993).

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences, and the like, for homologous recombination, as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome.

Suitable prokaryote selectable markers, useful for preparation of plant expression cassettes, include resistance toward antibiotics such as ampicillin, tetracycline, or kanamycin. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art. Usually, the plant selectable marker gene will encode antibiotic resistance, with suitable genes including at least one set of genes coding for resistance to the antibiotic spectinomycin, the streptomycin phosphotransferase (*spt*) gene coding for streptomycin resistance, the neomycin phosphotransferase (*nptII*) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (*hpt* or *aphiv*) gene encoding resistance to hygromycin, acetolactate synthase (*als*) genes and modifications encoding resistance to, in particular, the sulfonylurea-type herbicides, genes coding for resistance to herbicides which act to inhibit the action of glutamine synthase such as phosphinothricin or basta (e.g., the bar gene), or other similar genes known in the art.

The constructs of the subject invention will include the expression cassette for expression of the protein(s) of interest. Generally, there will be at least one expression cassette, and two or more are feasible, including a selection cassette. The recombinant expression cassette contains, in addition to the heterologous protein encoding sequence, at least the following elements: a promoter region, signal sequence, 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and transcription and translation termination sequences. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

In a preferred aspect of the present invention, a gene encoding a heterologous polypeptide, or a functional equivalent thereof, and/or a gene encoding a post-translational modification enzyme is inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence, or in the case of an RNA viral vector, the necessary elements for replication and translation. Methods for providing transgenic plants of the present invention comprise

constructing expression vectors containing a protein coding sequence, and expression vectors containing a sequence that encodes a post-translational enzyme or subunit thereof important to the biosynthesis of the protein, and/or an appropriate signal peptide coding sequence, and appropriate transcriptional/translational control signals. These methods include *in vitro* recombinant DNA techniques, synthetic techniques and *in vivo* recombination/genetic recombination. *See, e g.*, TRANSGENIC PLANTS: A PRODUCTION SYSTEM FOR INDUSTRIAL AND PHARMACEUTICAL PROTEINS (Owen & Pen eds., John Wiliey & Sons, 1996); GALUN & BREIMAN DES, TRANSGENIC PLANTS (Imperial College Press, 1997); APPLIED PLANT BIOTECHNOLOGY (Chopra, Malik, & Bhat eds., Science Publishers, Inc., 1999); U.S. Pat. Nos. 5,620,882; 5,959,177; 5,639,947; 5,202,422; and 4,956,282; WO 97/38710.

Heterologous Coding Sequences

Various pharmaceutically or commercially relevant polypeptides have been expressed in plant systems to varying degrees. The heterologous coding sequence may be for any protein of interest, either prokaryotic or eukaryotic, particularly eukaryotic. The gene providing the desired product will particularly be those genes associated with heterologous polypeptides and enzymes that are capable of performing post-translational modifications.

Specifically, the present invention may be used to produce post-translational modification enzymes, which include, oxidases, carboxylases, hydroxylases, prolyl 4-hydroxylase or a subunit thereof, lysyl oxidase, lysyl hydroxylase, C-proteinase, N-proteinase, PACE, γ-glutamyl carboxylase, N-acetylglucosaminaltransferases, N-acetlygalactosaminyltransferases, N-acetlygalactosaminyltransferases, sialyl-transferases, fucosyltransferases, galactosyltransferases, mannosyltransferases, sulfotransferases, glycosidases, acetyltransferases, and mannosidases.

Additionally, proteins of interest are mammalian proteins. Such proteins include, but are not limited to blood proteins (such as, serum albumin, Factor VII, Factor VIII (or modified Factor VIII), Factor IX, Factor X, tissue plasminogen factor, Protein C, von Willebrand factor, antithrombin III, and erythropoietin), colony stimulating factors (such as, granulocyte colony-stimulating factor (G-CSF), macrophage colony-stimulating factor (M-CSF), and granulocyte macrophage colony-stimulating factor (GM-CSF)), cytokines (such as, interleukins), integrins, addressins, selectins, homing receptors, surface membrane proteins (such as, surface membrane protein receptors), T cell receptor units, immunoglobulins,

soluble major histocompatibility complex antigens, structural proteins (such as, collagen, fibroin, elastin, tubulin, actin, and myosin), growth factor receptors, growth factors, growth hormones, cell cycle proteins, vaccines, fibrinogen, thrombin, cytokines, hyaluronic acid and antibodies.

Signal Sequence

Also included in chimeric genes used in the practice of the methods of the present invention are signal secretion sequences. In addition to encoding the protein of interest, the chimeric gene also encodes a signal peptide that allows processing and translocation of the protein, as appropriate. The chimeric gene typically lacks any sequence that might result in the binding of the desired protein to the appropriate membrane to achieve the proper post-translational modification(s). The signal sequences may be derived from plants such as wheat, barley, cotton, and rice. Mammalian or other non-plant signal sequences may also be useful, although it is typical to use a signal sequence from the same kingdom as the protein to be expressed. These signal sequences will direct the nascent polypeptide to the endoplasmic reticulum, where the polypeptide will subsequently undergo post-translational modification. One preferred aspect of the present invention uses the signal peptide derived from the tobacco extensin signal peptide.

Those of skill can routinely identify new signal peptides. Plant signal peptides typically have a tripartite structure, with positively-charged amino acids at the N-terminal end, followed by a hydrophobic region and then the cleavage site within a region of reduced hydrophobicity. The conservation of this mechanism is demonstrated by the fact that cereal α-amylase signal peptides are recognized and cleaved in foreign hosts such as *E. coli* and *S. cerevisiae*, however particular signal sequences may allow higher expression in some hosts.

The flexibility of this mechanism is reflected in the wide range of polypeptide sequences that can serve as signal peptides. Thus, the ability of a sequence to function as a signal peptide may not be evident from casual inspection of the amino acid sequence. Methods designed to predict signal peptide cleavage sites identify the correct site for only about 75% of the sequences analyzed. *See* Heijne Gv: Cleavage-site motifs in protein targeting sequences. In 14 GENETIC ENGINEERING (Setlow ed., Plenum Press, New York 1992).

Although, sequence homology is not always present in the signal peptides, hydrophilicity plots demonstrate that the signal peptides of these genes are relatively hydrophobic.

Promoters

To produce the desired protein expression in plants, the expression of the heterologous protein may be under the direction of a plant promoter. Promoters suitable for use in accordance with the present invention are described in the art. *See e.g.*, WO 91/198696. Examples of promoters that may be used in accordance with the present invention include non-constitutive promoters or constitutive promoters, such as, the nopaline synthetase and octopine synthetase promoters, cauliflower mosaic virus (CaMV) 19S and 35S promoters, and the figwort mosaic virus (FMV) 35 promoter. *See* U.S. Pat. No. 6,051,753.

In one aspect of the present invention, the heterologous protein or post-translational enzyme or nucleic acid affecting the natural post-translational modification abilities of the host may be expressed in a specific tissue, cell type, or under more precise environmental conditions or developmental control. Promoters directing expression in these instances are known as inducible promoters. In the case where a tissue-specific promoter is used, protein expression is particularly high in the tissue from which extraction of the protein is desired. Depending on the desired tissue, expression may be targeted to the endosperm, aleurone layer, embryo (or its parts as scutellum and cotyledons), pericarp, stem, leaves, tubers, roots, etc. Examples of known tissue-specific promoters include the tuber-directed class I patatin promoter, the promoters associated with potato tuber ADPGPP genes, the soybean promoter of beta-conglycinin (7S protein) which drives seed-directed transcription, and seed-directed promoters such as those from the zein genes of maize endosperm and rice glutelin-1 promoter. *See, e.g.*, Bevan et al., 14 NUCLEIC ACIDS RES. 4625-38 (1986); Muller et al., 224 MOL. GEN. GENET. 136-46 (1990); Bray, 172 PLANTA 364-70 (1987); Pedersen et al., 29 CELL 1015-26 (1982); Russell & Fromm, 6 TRANSGENIC RES. 157-58 (1997).

In a preferred aspect of the invention, the heterologous protein or post-translational modifying enzyme or nucleic acid affecting the natural post-translational modification abilities of the host is produced from seed by way of seed-based production techniques using, for example, canola, corn, soybeans, rice and barley seed. *See, e.g.*, Russell, 240 CURRENT

TECHNOLOGIES IN MICROBIOL. & IMMUNOL. 119-38 (1999). In such a process, the desired protein is recovered during or after seed maturation, or during the germination phase.

In yet another aspect of the present invention, promoters may be used to direct the expression of antisense nucleic acids to reduce, increase, or alter the expression level of plant-specific post-translational modification enzymes and other plant-specific proteins in a desired tissue.

Transcription and Translation Terminators

The expression cassettes or chimeric genes of the present invention typically have a transcriptional termination region at the opposite end from the transcription initiation regulatory region. The transcriptional termination region may normally be associated with the transcriptional initiation region or from a different gene. The transcriptional termination region may be selected, particularly for stability of the mRNA to enhance expression. Illustrative transcriptional termination regions include the NOS terminator from *Agrobacterium* Ti plasmid and the rice α-amylase terminator.

Polyadenylation tails are also commonly added to the expression cassette to optimize high levels of transcription and proper transcription termination, respectively. Alber and Kawasaki, 1 MOL. AND APPL. GENET. 419-34 (1982). Polyadenylation sequences include but are not limited to the *Agrobacterium* octopine synthetase signal, (Gielen, et al., 3 EMBO J. 835-46 (1984)), or the nopaline synthase of the same species (Depicker, et al., 1 MOL. APPL. GENET. 561-73 (1982)).

Enhancers

Enhancer sites, which are standard and known to those in the art, may be included in the expression cassettes to increase and/or maximize transcription of the heterologous protein or post-translational enzyme or nucleic acid affecting the natural post-translational modification abilities of the host in a plant host system. These include, but are not limited to, peptide export signal sequence, codon usage, introns, polyadenylation, and transcription termination sites. Methods of modifying nucleic acid constructs to increase expression levels in plants are also generally known in the art. *See, e.g.* Rogers et al., 260 J. BIOL. CHEM. 3731-38 (1985); Cornejo et al., 23 PLANT MOL. BIOL. 567-81 (1993).

In engineering a plant system that affects the rate of transcription of a heterologous protein or post-translational enzyme or nucleic acid affecting the natural post-translational modification abilities of the host, various factors known in the art including regulatory sequences such as positively or negatively acting sequences, enhancers and silencers, as well as, chromatin structure can affect the rate of transcription in plants. The present invention provides that at least one of these factors may be utilized in engineering plants to express a heterologous protein and post-translational enzymes.

Transformation of Plant Cells

Vectors containing a chimeric gene of the present invention can be introduced into a plant host system by a variety of techniques. In a preferred aspect of the present invention, a first vector comprising a heterologous protein and a second vector comprising a post-translational modification enzyme or nucleic acid affecting the natural post-translational modification abilities of the host may be co-introduced into the plant host system. In another preferred aspect of the present invention, a multi-cistronic vector comprising a heterologous protein and the post-translational modification enzyme or nucleic acid affecting the natural post-translational modification abilities of the host may be introduced into the plant cell using the methods described below.

These vectors may include selectable markers for use in plant cells (such as, the *nptII* kanamycin resistance gene). The vectors may also include sequences that allow their selection and propagation in a secondary host, such as, sequences containing an origin of replication and a selectable marker. Typical secondary hosts include bacteria and yeast.

The vectors of the present invention may also be modified to intermediate plant transformation plasmids that contain a region of homology to an *Agrobacterium tumefaciens* vector, a T-DNA border region from *A. tumefaciens*, and chimeric genes or expression cassettes. Further, the vectors of the invention may comprise a disarmed plant tumor inducing plasmid of *A. tumefaciens*. However, few monocots appear to be natural hosts for *Agrobacterium*. Although transgenic plants have been produced in gladiolus using *Agrobacterium* vectors as described by Graves & Goldman, 169 J. BACTERIOL. 1745-46 (1987); Ishida et al., 14 NAT. BIOTECH. 745-50 (1996). Therefore, commercially important cereal grains such as rice, corn, and wheat must be transformed using alternative methods.

Standard methods for the transformation of rice, wheat, corn, sorghum, and barley are described in the art. *See* Christou et al., 10 TRENDS IN BIOTECHNOLOGY 239 (1992); Lee et al., 88 P.N.A.S. 6389-93 (1991). Wheat can be transformed by techniques similar to those employed for transforming corn or rice. Furthermore, Casas et al., 90 P.N.A.S. 11212-16 (1993), describe a method for transforming sorghum, while Lazzeri, 49 METHODS MOL. BIOL. 95-106 (1995), teach a method for transforming barley. Suitable methods for corn transformation are provided by Fromm et al., 8 BIO/TECHNOLOGY 833-39 (1990); Gordon-Kamm et al., 2 PLANT CELL 603-18 (1990); Russell et al., 6 TRANSGENIC RES., 157-58 (1997); U.S. Pat. No. 5,780,708.

Vectors useful in the practice of the present invention may be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, 202 MOL. GEN. GENET., 179-85 (1985). The genetic material may also be transferred into the plant cell by using polyethylene glycol, Krens et al., 96 NATURE 72-74 (1982).

Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface. Klein et al., 327 NATURE 70-73 (1987); Knudsen & Muller, 185 PLANTA 330-36 (1991).

Additionally, another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley et al., 79 P.N.A.S. 1859-63 (1982).

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., 82 P.N.A.S. 5824-28 (1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus. *See* U.S. Pat. No. 5,584,807.

Inserting A Gene Coding For A Second Polypeptide Into A Plant Species

Useful genes include those genes coding for a second polypeptide that can autogenously associate with the first polypeptide in such a way as to form a biologically active protein, or post-translationally modify a target heterologous protein. The methods

used to introduce a gene coding for this second polypeptide into a member of a plant species are the same as the methods used to introduce a gene into the first member of the same plant species and have been described above. These include introducing both of the polypeptides on a single vector with appropriate regulatory control elements, co-transforming a single plant host tissue with separate vectors, each of which express a first or second polypeptide; introducing a second polypeptide into a plant host system that already expresses a first polypeptide; and generating two separate plants that each express different heterologous polypeptides, and then sexually crossing them to generate a plant that expresses both heterologous polypeptides. *See* WO 98/10,062; Mendel, EXPERIMENTS IN PLANT HYBRIDIZATION, (Oliver Boyd ed., Edinburgh, Scotland, 1965) (1865); U.S. Pat No. 5,202,422.

Vector Transformation

A common vector method of introducing the vector into plant cells is to infect a plant cell with *Agrobacterium tumefaciens* previously transformed with the gene. Under appropriate conditions known in the art, the transformed plant cells are grown to form shoots or roots, and develop further into plants.

Agrobacterium is a representative genus of the gram-negative family *Rhizobiaceae*. Its species are responsible for plant tumors such as crown gall and hairy root disease. In the dedifferentiated tissue characteristic of the tumors, amino acid derivatives known as opines are produced and catabolized. The bacterial genes responsible for expression of opines are a convenient source of control elements for chimeric expression cassettes.

Heterologous genetic sequences, such as the chimeric genes of the present invention, can be introduced into appropriate plant cells, by means of the Ti plasmid of *A. tumefaciens*. The Ti plasmid is transmitted to plant cells on infection by *A. tumefaciens*, and is stably integrated into the plant genome. Schell, 237 SCIENCE 1176-83 (1987).

Ti plasmids contain two regions essential for the production of transformed cells. One of these, named transferred DNA (T-DNA), is transferred to plant nuclei and induces tumor formation. The other, termed virulence region, is essential for the transfer of this T-DNA but is not itself transferred. The transferred DNA region, which transfers to the plant genome, can be increased in size by the insertion of the gene encoding group 3 LEA proteins without its ability to be transferred being affected.

Promoters directing expression of selectable markers used for plant transformation (e.g., *nptII*) should operate effectively in plant hosts. One such promoter is the nos promoter from native Ti plasmids. Herrera-Estrella et al., 303 NATURE 209-13 (1983). Others include the 35S and 19S promoters of cauliflower mosaic virus, Odell et al., 313 NATURE 810-12 (1985), and the 2' promoter, Velten et al., 3 EMBO J. 2723-30 (1984).

Isolating Progeny Containing The Multimeric Protein

Progeny containing the desired heterologous protein or post-translational modification enzyme or nucleic acid affecting the natural post-translational modification abilities of the host can be identified by assaying for the presence of the biologically active heterologous protein using assay methods well known in the art. Such methods include Western blotting, immunoassays, binding assays, and any assay designed to detect a biologically functional heterologous protein. See, for example, the assays described in KLEIN, IMMUNOLOGY: THE SCIENCE OF SELF-NONSELF DISCRIMINATION (John Wiley & Sons eds., New York, N.Y. 1982).

Preferred screening assays are those where the biologically active site on the heterologous protein is detected in such a way as to produce a detectable signal. This signal may be produced directly or indirectly and such signals include, for example, the production of a complex, formation of a catalytic reaction product, the release or uptake of energy, and the like. For example, a progeny containing an antibody molecule produced by this method may be processed in such a way to allow that antibody to bind its antigen in a standard immunoassay such as an ELISA or a radio-immunoassay similar to the immunoassays described in ANTIBODIES: A LABORATORY MANUAL (Harlow & Lane, eds., Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1988).

A further aspect of the present invention is a method of producing a heterologous protein that has been correctly processed to result in a biologically active protein. Generally, the method combines the elements of cultivating a plant of the present invention, and harvesting the plant that was cultivated to produce the desired heterologous protein.

A plant of the present invention containing the desired processed heterologous protein comprised of a first polypeptide encoding the heterologous protein and a second polypeptide encoding the post-translational modification enzyme is cultivated using methods well known

to one skilled in the art. Any of the transgenic plants of the present invention may be cultivated to isolate the desired heterologous protein they contain.

After cultivation, the transgenic plant is harvested to recover the produced heterologous protein. This harvesting step may consist of harvesting the entire plant, or only the leaves, or roots of the plant. This step may either kill the plant or if only the portion of the transgenic plant is harvested may allow the remainder of the plant to continue to grow.

Plant Regeneration

After determination of the presence and expression of the desired gene products, whole plant regeneration is desired. Plant regeneration from cultured protoplasts is described in EVANS, ET AL., HANDBOOK OF PLANT CELL CULTURES, Vol. 1: (MacMillan Publishing Co. New York 1983); CELL CULTURE & SOMATIC CELL GENETICS OF PLANTS, (Vasil I.R., ed., Acad. Press, Orlando, Vol. I 1984, and Vol. III 1986).

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables, dicots, and monocots.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted.

Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

The transgenic plants according to this invention can be used to develop hybrids or novel varieties embodying the desired traits. Such plants would be developed using traditional selection type breeding.

The mature plants, grown from the transformed plant cells, are selfed and non-segregating, homozygous transgenic plants are identified. An outcross can also be done, to move the gene into another plant. The transgenic plant produces seed containing the proteins of the present invention. The transgenic plants according to this invention can be used to develop hybrids or novel varieties embodying the desired traits. Such plants would be developed using traditional selection type breeding.

Antisense Applications

In addition to the above indicated genes, one may also have constructs which provide for inactivation of endogenously expressed genes. *See*, U.S. Pat. No. 6,015,939. Of particular interest is the inactivation of genes that are responsible for generating plant-specific post-translational modifications. These genes may include one or more of the amylases, e.g., RAmy3B, RAmy3C, RAmy3E or homologs thereof, glucosidases, and glycosyl transferases such as fucosyl transferase, mannosidase I, and mannosidase II. These genes are responsible for creating plant specific linkages that include, but are not limited to, β(1,2)-xylose and α(1,3)-fucose.

Inactivation of gene expression may be achieved in a number of ways. The most convenient is the use of an antisense sequence, where the antisense sequence may be complementary to any portion of the mRNA, including both the non-coding and coding regions. Normally, the antisense sequence will be at least about 30 nucleotides, more usually at least about 50 nucleotides, and may be up to or greater than the sequence of the mRNA to which the antisense sequence is complementary.

In one embodiment, the 3'-terminal sequence of the antisense sequence is selected to provide for mRNA stability, there being a number of sequences which are known to destabilize the mRNA which can be avoided.

The transcription initiation region for the antisense sequence may be constitutive or inducible. A relatively strong promoter may be employed, such as, the 35S CaMV promoter, the RUBSICO promoter, or beta-conglycinin promoter. To enhance the transcription of the antisense sequence, one may use various enhancers associated with other promoters to

increase the rate of transcription of the antisense sequence. Enhancers which find use include the 35S CaMV enhancer, and the introns of the alcohol dehydrogenase gene of maize.

Purification of proteins

The expressed heterologous protein of the invention may be purified to homogeneity by chromatography. In one embodiment, recombinant polypeptide is purified by size exclusion chromatography. However, other purification techniques known in the art can also be used, including ion exchange chromatography, and reverse-phase chromatography and selective phase separation. *See, e.g.*, MANIATIS et al., MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor Laboratory, N.Y. 1989); AUSUBEL et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Greene Publishing Associates and Wiley Interscience, N.Y. 1989); SCOPES, PROTEIN PURIFICATION: PRINCIPLES AND PRACTICE (Springer-Verlag New York, Inc., NY 1994), and U.S. Pat Nos. 5,990,284, 5,804694, and 6,037,456.

Nucleic Acid Sequences

In accordance with the invention, polynucleotide sequences which encode heterologous polypeptides or post-translational enzymes or functional equivalents thereof or nucleic acid affecting the natural post-translational modification abilities of the host, may be used to generate recombinant nucleic acid sequences that direct the expression of such proteins, or functional equivalents thereof, in plant cells.

Antibodies are a preferred form of heterologous polypeptide produced according to the invention. Antibodies, or immunoglobulins (Ig), are divided into several categories: IgG, IgM, IgE, IgD, and IgA. *See generally* Coleman et al., FUND. IMMUNOL. 79-84 (2d ed. 1992); Kuby, IMMUNOL. 124-29 (2d ed. 1994); Paul, FUND. IMMUNOL. 135-47 (1984); Roitt et al., IMMUNOL. 5.2-5.6 (2d ed. 1989).

As described above, fragments of antibodies are also well known in the art. For example, murine monoclonal antibody L6, of class IgG2, has been sequenced and recognizes antigens associated with human lung adenocarcinoma. Hellstrom et al., 46 CANCER RES. 3917-23 (1986); U.S. Pat. Nos. 4,906,562; 4,935,495; and 5,091,177. The L6 sFv single chain has been produced in high levels in tobacco. U.S. Pat. No. 6,080,560.

In another aspect of the present invention, polynucleotide sequences which encode post-translational enzymes or functional equivalents thereof or nucleic acid affecting the natural post-translational modification abilities of the host, may be used to generate recombinant DNA molecules that direct the expression of such post-translational enzymes or functional equivalents thereof or nucleic acid affecting the natural post-translational modification abilities of the host, in plant cells. The nucleic acids encoding many members of the glycosyltransferase enzyme family have been cloned and sequenced. For example, UDP-N-acetylglucosamine β-D-mannoside β-1,4 N-acetylglucosaminyltransferase III (GnT-III), an enzyme that transfers a GlcNAc residue in UDP-N-acetylglucosamine (UDP-GlcNAc) to a mannose (Man) residue forming a β 1-4 bond in an asparagine binding type sugar chain, has been cloned and sequenced. U.S. Pat. No. 5,874,271.

Further examples of sequenced human glycosyltransferases include human β-1,4 galactosyltransferase (Japanese Pat. No. 2027987) and variants of human galactosyltransferase (U.S. Pat. No. 5,955,282); human GDP-Fuc β-D-galactoside α (1,2) fucosyltransferase (U.S. Pat. No. 5,955,347); β 1,6-N-acetylglucosaminyltransferase (U.S. Pat. No. 5,766,713); and human β 1,6 N-acetylglucosaminyltransferase, which forms core 2 oligosaccharide structures in O-glycans (U.S. Pat. No. 5,360,733).

Mammalian glycosyltransferases have also been cloned and sequenced. Examples include mouse UDP-Gal β-D-Gal-(1,4)-D-GlcNac α (1,3) galactosyltransferase (U.S. Pat. No. 5,955,347), and UDP-N-acetylglucosamine α 6-D-mannoside β 1,6-N-acetylglucosaminyltransferase (GlcNAc T-V) (U.S. Pat. No. 5,602,003).

Additionally, a β-galactoside α 2,6 sialyltransferase has been cloned into mammalian CHO cells, and expressed in addition to the CHO-native 1,3 transferase. (U.S. Pat. No. 5,047,335). This patent does not, however, address the challenge of expressing such a gene in plants.

The DNA coding for other post-translational modifying enzymes involved in glycosylation have also been identified. For example, the sulfotransferase, chondrotin 6-sulfotransferase (C6ST) which transfers a sulfate group to the hydroxyl group at C-6 position of N-acetylgalactosamine residue or galactose residue of glycosaminoglycan is provided by Habuchi et al., U.S. Pat. No. 5,827,713.

In another aspect of the present invention, polynucleotide sequences which encode a heterologous polypeptide or post-translational enzymes or functional equivalents thereof or nucleic acid affecting the natural post-translational modification abilities of the host, may be used to generate recombinant DNA molecules that direct the expression of such post-translational enzymes or heterologous polypeptide, or functional equivalents thereof or nucleic acid affecting the natural post-translational modification abilities of the host, in plant cells.

In another preferred embodiment, such polynucleotide sequences encode post-translational enzymes, as described above, or functional equivalents thereof or nucleic acid affecting the natural post-translational modification abilities of the host. For example, the nucleic acid sequences encoding the alpha 2 subunit of prolyl 4-hydroxylase have been generally described in the art. U.S. Pat. No. 5,928,922. In a preferred embodiment of the invention, such polynucleotide sequences may include sequences that encode collagen types I through XX as described above or functional equivalents thereof. The nucleic acid sequences encoding the known collagen types have been generally described in the art. *See, e.g.*, Fukai et al., 245 METH. ENZYMOL. 3-28 (1994), and references cited therein; WO 97/38,710. Ovine, chicken and other animal collagens have also been described in literature easily accessible to those of ordinary skill in the art. *See, e.g.*, Turner et al., 89(2) J. CLIN. INVEST. 592-601 (1992), Rhode et al., 179(3) BIOCHEM. J. 631-42 (1979).

Post-translational enzymes are involved in the production of collagen. As naturally produced, collagens are structural proteins comprised of one or more collagen subunits which together form at least one triple-helical domain. A variety of enzymes are utilized in order to transform the collagen subunits into procollagen or other precursor molecules and then mature collagen. Such enzymes include prolyl 4-hydroxylase, C-proteinase, N-proteinase, lysyl oxidase and lysyl hydroxylase.

Prolyl 4-hydroxylase is a $\alpha_2\beta_2$ tetramer, and plays a central role in the biosynthesis of all collagens, as the 4-hydroxyproline residues stabilize the folding of the newly synthesized polypeptide chains into triple-helical molecules. *See*, Prockop et al., 64 ANNU. REV. BIOCHEM. 403-34 (1995); Kivirikko et al., POST-TRANSLATIONAL MODIFICATIONS OF PROTEINS (Harding & Crabbe eds., CRC Press, Boca Raton, FL 1992); Kivirikko et al., 3 FASEB J. 1609-17 (1989); U.S. Pat. Nos. 5,928,922; 5,593,859; and 5,405,757. A human isoform of prolyl 4-hydroxylase has been cloned and characterized as described in

Helaakoski et al., 92 P.N.A.S. 4427-31 (1995). The nucleic acid sequences encoding the α2 subunit of prolyl 4-hydroxylase have been generally described in the art. *See, e.g.,* U.S. Pat. No. 5,928,922.

Lysyl hydroxylase, an α2 homodimer, catalyzes the post-translation modification of collagen to form hydroxylysine in collagens. *See generally*, KIVIRIKKO et al., POST-TRANSLATIONAL MODIFICATIONS OF PROTEINS, *supra*; KIVIRIKKO, PRINCIPLES OF MEDICAL BIOLOGY. VOL. 3: CELLULAR ORGANELLES AND THE EXTRACELLULAR MATRIX (Bittar & Bittar eds., JAI Press, Greenwich, Great Britain 1995). Isoforms of lysyl hydroxylase have been cloned and identified. *See, e.g.* Passoja et al., 95(18) P.N.A.S. 10482-86 (1998); Valtavaara et al., 272(11) J. BIOL. CHEM. 6831-34 (1997).

Lysyl oxidase is an extracellular copper enzyme that catalyzes the oxidative deamination of the α-amino group in certain lysine and hydroxylysine residues to form a reactive aldehyde. These aldehydes then undergo an aldol condensation to form aldols, which crosslink collagen fibrils. Information on the DNA and protein sequence of lysyl oxidase can found, among elsewhere, in Kivirikko, PRINCIPLES OF MEDICAL BIOLOGY, *supra*; Kagan, 190 PATH. RES. PRACT. 910-19 (1994); Kenyon et al., 268(25) J. BIOL. CHEM. 18435-37 (1993), Woo et al., 267(34) J. BIOL. CHEM. 24199-06 (1992); Mariani et al., 12(3) MATRIX 242-48 (1992); and Harnalainen et al., 11(3) GENOMICS 508-16 (1991).

The nucleic acid sequences encoding a number of other post-translational enzymes have been reported. *See, e.g.* Vuori et al., 89 P.N.A.S. 7467-70 (1992); Kessler et al., 271 SCIENCE 360-62 (1996). The nucleic acid sequences encoding the various post-translational enzymes may also be determined according to the methods generally described above and include use of appropriate probes and nucleic acid libraries.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding post-translational enzymes and proteins, or functional equivalents thereof, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code.

New post-translational enzymes or heterologous polypeptides or known post-translational enzymes or heterologous polypeptides or nucleic acid affecting the natural post-

translational modification abilities of the host for which nucleic acid sequences are not available may be obtained from cDNA libraries prepared from tissues believed to possess a "novel" type of post-translational enzyme or heterologous polypeptide and to express the novel post-translational enzymes or heterologous polypeptide at a detectable level. For example, a cDNA library could be constructed by obtaining polyadenylated mRNA from a cell line known to express the novel heterologous polypeptide, or a cDNA library previously made to the tissue/cell type could be used. The cDNA library is screened with appropriate nucleic acid probes, and/or the library is screened with suitable polyclonal or monoclonal antibodies that specifically recognize other heterologous polypeptides. Appropriate nucleic acid probes include oligonucleotide probes that encode known portions of the novel post-translational enzymes or collagen from the same or different species. Other suitable probes include, without limitation, oligonucleotides, cDNAs, or fragments thereof that encode the same or similar gene, and/or homologous genomic DNAs or fragments thereof. Screening the cDNA or genomic library with the selected probe may be accomplished using standard procedures known to those in the art, such as those described in Chapters 10-12 of SAMBROOK et al., MOLECULAR CLONING: A LABORATORY MANUAL (Cold Spring Harbor Laboratory Press, New York, 1989). Other means for identifying novel heterologous polypeptides or post-translational enzymes may involve known techniques of recombinant DNA technology, such as by direct expression cloning or using the polymerase chain reaction (PCR) as described in U.S. Pat. No. 4,683,195, or in Chapter 14 of SAMBROOK et al., *supra*, or in Chapter 15 of CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (Ausubel et al., eds., Gireene Publishing Associates and Wiley-Interscience 1991).

Altered DNA sequences which may be used in accordance with the invention include deletions, additions or substitutions of different nucleotide residues resulting in a sequence that encodes the same or a functionally equivalent gene product. The gene product itself may contain deletions, additions or substitutions of amino acid residues within a post-translational enzyme or heterologous polypeptide sequence, which result in a functionally equivalent post-translational enzyme or heterologous protein.

The nucleic acid sequences of the invention may be engineered in order to alter the coding sequence for a variety of ends including, but not limited to, alterations that modify expression and processing of the gene product. For example, alternative secretory signals may be substituted for the native secretory signal (e.g., U.S. Pat. No. 5,716,802). More

specifically, the KDEL sequence has been shown to increase the expression of single-chain antibody in tobacco. Schouten et al., 30(4) PLANT MOL. BIOL. 781-93 (1996). Additional mutations may be introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, or alter glycosylation or phosphorylation patterns.

Additionally, when expressing in non-human cells, the polynucleotides encoding the post-translational enzymes or heterologous proteins or nucleic acid affecting the natural post-translational modification abilities of the host of the invention may be modified in the silent position of any triplet amino acid codon so as to better conform to the codon preference of the particular host organism. More specifically, translational efficiency of a protein in a given host organism can be regulated through codon bias, meaning that the available 61 codons for a total of 20 amino acids are not evenly used in translation, and observation that has been made for prokaryotes (Kane, 6 CURRENT OP. BIOTECH. 494-500 (1995)), and eukaryotes (ERNST, CODON USAGE & GENE EXPRESSION 196-99 (Elsevier Pub., Cambridge 1988). An application of these observations, i.e., the adaptation of the codon bias of a bacterial gene to the codon bias of a higher plant, resulted in significantly higher accumulation of the foreign protein in the plant. Perlak et al., 88(8) P.N.A.S. 3324-28 (1991); *see also* Murray et al., 17 NUCL. ACIDS RES. 477-98 (1989). Codon usage tables have been established not only for organisms, but also for organelles and specific tissues (Kazusa DNA Research Institute, <www.kazusa.or.jp>), and their general availability enables researchers to adopt the codon usage of a given gene to the host organism. Other factors like the context of the initiator methionine start codon (Kozak, 234 GENE 187-208 (1999)), may influence the translation rate of a given protein in a host organism, and can therefore be taken into consideration. *See also* Taylor et al., 210 MOL. GENETICS 572-77 (1987). Translation may also be optimized by reference to codon sequences that may generate potential signals of intron splice sites. PLANT MOLECULAR BIOLOGY LABFAX (Croy, ed. 1993), mRNA instability and polyadenylation signals (Perlak et al., *supra*).

The nucleic acid sequences of the invention are further directed to sequences that encode variants of the described post-translational enzymes and heterologous proteins and fragments or nucleic acid affecting the natural post-translational modification abilities of the host. These amino acid sequence variants of native heterologous proteins or post-translational enzymes and fragments or nucleic acid affecting the natural post-translational

modification abilities of the host may be prepared by methods known in the art by introducing appropriate nucleotide changes into a native or variant heterologous protein or post-translational enzyme encoding polynucleotide. There are two variables in the construction of amino acid sequence variants: the location of the mutation and the nature of the mutation. The amino acid sequence variants are preferably constructed by mutating the polynucleotide to give an amino acid sequence that does not occur in nature. These amino acid alterations can be made at sites that differ in post-translational enzymes or heterologous protein, from different species (variable positions) or in highly conserved regions (constant regions). Sites at such locations will typically be modified in series, e.g., by substituting first with conservative choices (e.g., hydrophobic amino acid to a different hydrophobic amino acid) and then with more distant choices (e.g., hydrophobic amino acid to a charged amino acid), and then deletions or insertions may be made at the target site.

Amino acids are divided into groups based on the properties of their side chains (polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature): (1) hydrophobic (leu, met, ala, ile); (2) neutral hydrophobic (cys, ser, thr); (3) acidic (asp, glu); (4) weakly basic (asn, gln, his); (5) strongly basic (lys, arg); (6) residues that influence chain orientation (gly, pro); and (7) aromatic (trp, tyr, phe). Conservative changes encompass variants of an amino acid position that are within the same group as the native amino acid. Moderately conservative changes encompass variants of an amino acid position that are in a group that is closely related to the native amino acid (e.g., neutral hydrophobic to weakly basic). Non-conservative changes encompass variants of an amino acid position that are in a group that is distantly related to the "native" amino acid (e.g., hydrophobic to strongly basic or acidic).

Amino acid sequence deletions generally may range from about 1 to 30 residues, preferably about 1 to 10 residues, and are typically contiguous. Amino acid insertions include amino- and/or carboxyl-terminal fusions ranging in length from one to one hundred or more residues, as well as intrasequence insertions of single or multiple amino acid residues. Intrasequence insertions may range generally from about 1 to 10 amino residues, preferably from 1 to 5 residues. Examples of terminal insertions include the heterologous signal sequences necessary for secretion or for intracellular targeting in different host cells.

In one method, polynucleotides encoding a heterologous polypeptide or a post-translational enzyme or nucleic acid affecting the natural post-translational modification

abilities of the host are changed via site-directed mutagenesis. This method uses oligonucleotide sequences that encode the polynucleotide sequence of the desired amino acid variant, as well as a sufficient adjacent nucleotide on both sides of the changed amino acid to form a stable duplex on either side of the site of being changed. In general, the techniques of site-directed mutagenesis are well known to those of skill in the art and this technique is exemplified by publications such as, Adelman et al., 2 DNA 183-93 (1983). A versatile and efficient method for producing site-specific changes in a polynucleotide sequence was published by Zoller & Smith, 10 NUCLEIC ACIDS RES. 6487-500 (1982).

Mutations provide one or more unique restriction sites and do not alter the amino acid sequence encoded by the nucleic acid molecule, but merely provide unique restriction sites useful for manipulation of the molecule. Thus, the modified molecule would be made up of a number of discrete regions, or D-regions, flanked by unique restriction sites. These discrete regions of the molecule are herein referred to as cassettes. Molecules formed of multiple copies of a cassette are another variant of the present gene which is encompassed by the present invention. Recombinant or mutant nucleic acid molecules or cassettes which provide desired characteristics such as resistance to endogenous enzymes such as collagenase are also encompassed by the present invention.

PCR may also be used to create amino acid sequence variants of a post-translational enzyme or heterologous polypeptide or nucleic acid affecting the natural post-translational modification abilities of the host. When small amounts of template DNA are used as starting material, primer(s) that differs slightly in sequence from the corresponding region in the template DNA can generate the desired amino acid variant. PCR amplification results in a population of product DNA fragments that differ from the polynucleotide template encoding the heterologous polypeptide at the position specified by the primer. The product DNA fragments replace the corresponding region in the plasmid and this gives the desired amino acid variant.

A further technique for generating amino acid variants is the cassette mutagenesis technique described in Wells et al., 34 GENE 315 (1985); and other mutagenesis techniques well known in the art, such as, for example, the techniques in Sambrook et al., *supra*; CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, Ausubel et al., *supra*.

Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence or

polypeptide, specifically, comprising a consistent (Gly-X-Y), amino acid structure, that are natural, synthetic, semi-synthetic, or -recombinant, may be used in the practice of the claimed invention. Such DNA sequences may be include those which are capable of hybridizing to the appropriate heterologous polypeptide sequence under stringent conditions.

EXAMPLES

Without further elaboration, it is believed that one skilled in the art, using the preceding description, can utilize the present invention to the fullest extent. The following examples are illustrative only, and not limiting of the remainder of the disclosure in any way whatsoever.

Example 1: Mammalian Glycosylation of Proteins in Plant Tissues

Recent advances in genetic engineering has enabled the use of plants for the production of therapeutic proteins, particularly monoclonal antibodies (Mabs). The IgG class of antibodies is a major protective immunoglobulin class, and thus, a major target for therapeutic design. Currently, the commercially available Mabs are in the IgG class. In general, the heavy chain of an IgG molecule has a conserved N-glycosylation site that is present in the constant region (C_H2) of the heavy chain. The glycan chains of each heavy chain appear to fill the interstitial region between the two C_H2 domains, without influencing the protein tertiary and quaternary structures which are essential for epitope binding. Removal of the glycan chains on an IgG molecule (aglycosylation) by genetic manipulation or by chemical inhibition resulted in decreased ADCC and CMC activity and increased *in vitro* sensitivity to proteases (Wright & Morrison, 1997). These data strongly suggest that N-glycosylation is critical for the full functionality of therapeutic antibodies.

For proper protein synthesis, the heavy and light chain peptides of an IgG antibody initially include N-terminal signal peptides that target the molecules to the ER. Subsequently, the signal peptides are cleaved, and assembly of heavy and light chains occurs through disulfide bond formation. Concurrently, a core glycan chain from dolichol phosphate is transferred to the asparagine residue of the heavy chain (Figures 1 and 16). Trimming of the mannose moieties and extension of the core structure by the addition of N-acetyl glucosamine (GlcNac), galactose, and terminal sialic acid (complex glycan) is catalyzed by glycosidases and glycosyltransferases located in the Golgi apparatus. The glycosylation

pattern at any one site can vary depending on the particular protein structure, cell type, or cellular compartment.

Initially, N-glycosylation patterns are similar in all eukaryotes, but in the later stages of complex glycan formation, the patterns of plants, insects, and yeast begin to differ from the mammalian systems because several of the glycosyltransferases present in the mammalian Golgi systems are absent in plants, insects and yeast. For example, a mammalian glycoprotein may include a terminal galactose and sialic acid as well as an α(1,6)-fucose at the proximal GlcNAc (Figure 2). However, in plants, neither sialic acid nor the same galactose linkage are found at the termini of complex type glycans. Instead, proteins tend to accumulate in plants without the terminal GlcNAc, and novel sugar groups are later added including β(1,2)-xylose linked to the β-mannose and an α(1,3)-fucose residue linked to the proximal glucosamine (Lerouge et al., 1998; Cabanes-Macheteau et al., 1999). Several studies have suggested that plant glycoproteins contain highly immunogenic glycans and the plant specific sugars β(1,2)-xylose and α(1,3)-fucose may be responsible for this immunogenic reaction.

To minimize the potential of immunogenic/allergic reaction of plant specific sugars and to ensure complete functionality of plant-generated immunoglobulins, it would be beneficial to establish mammalian glycosylation in plants. The present example thus relates to the expression of mouse, bovine, and human galactosyltransferase (GalT; β1,4 galactosyltransferase; EC 2.4.1.22) in tobacco and corn cell cultures. Additionally, based on the studies done by Palacpac et al., (1999), the addition of galactose at the terminal end of the glycan chain may limit the addition of β(1,2)-xylose and α(1,3)-fucose residues, and thereby enhance mammalian-type glycosylation in plants.

Procedures

Gene-specific PCR primers were designed using the MacVector program and the following published sequences that encode β-1,4 galactosyltransferase: a) mouse galactosyltransferase (GenBank accession number J03880; Shaper et al., 263 J. BIOL. CHEM. 10420-28 (1988)); b) bovine galactosyltransferase (GenBank accession number X14558; D'Agostaro et al., 183 EUR. J. BIOCHEM. 211-17 (1989)); and c) human galactosyltransferase (GenBank accession number M22921; Masri et al., 157 BIOCHEM. BIOPHYS. RES. COMMUN.

657-63 (1988)) (Figure 4). The primers were synthesized by Life Technologies (Rockville, MD).

Poly A^+ RNA from mouse, bovine, and human liver tissues (Clontech Laboratories, Inc., Palo Alto, CA) was incubated with the gene specific 3' primer for first strand cDNA synthesis according to the manufacturers' protocol provided in the RT-PCR kit (Life Technologies, Rockville, MD). The reaction products were later used for the amplification of cDNAs using the standard PCR reaction conditions. The PCR reaction mix contained 1x High Fidelity PCR buffer (Boehringer Mannheim, Indianapolis, IN), 10 mM dNTP, 50 mM $MgSO_4$, 20 pmole of 5' primer and 3' primer, 2.0 µl of the first strand cDNA, and 0.5 µl of Taq polymerase. The PCR reaction cycle was as follows: for the first 5 cycles, 94°C for 30s, 65°C for 1 minute, and 72°C for 2 minutes; for an additional 5 cycles, 94°C for 30s, 62°C for 1 minute, and 72° C for 2 minutes; and for the final 25 cycles, 94°C for 30s, 55°C for 1 minute, and 72°C for 2 minutes. After the reaction, 20 µl of the reaction products were separated on 1.0% agarose gel. As shown in Figure 5, the size of the bands amplified in each reaction correlates with the expected size. The correct size fragments were eluted from the gel and purified using Promega Wizard DNA purification kit (Promega, Madison, WI). The purified fragments were used for sub-cloning into various pMON vectors.

The PCR amplified cDNAs of mouse, bovine, and human GalT were first subcloned into pMON19469 containing the ampicillin backbone for transient expression analysis and DNA sequencing.

Following expression analysis, a cassette containing 35S CaMV promoter/HSP70intron/ mouse, bovine, or human GalT cDNA/NOS-3,' was subcloned into pMON41410 (derived from the modification of pMON25488) that has a CP4 selection marker gene. The details of the design of the vectors used to transform corn (pMON41411, pMON41412, and pMON41413) are illustrated in Figure 6. In addition to the GalT constructs, the corn cell lines were also co-bombarded with vectors that encode heavy and light chains of huNR-LU-10 monoclonal antibody. The construction of the vector containing the cDNAs of heavy and light chains is described in Figure 6.

In the plasmids used to transform tobacco, the HSP70 was replaced with a dssu linker (DAR298: 5'-GAT CTT CTA AGA AGA AGA ACT-3' and DAR299: 5'-CTA GAG TTC TTC TTC TTA GAA-3') as a 5' leader sequence. The entire cassette containing 35S CaMV

promoter/dssu linker/ mouse, bovine, or human GalT cDNA/NOS-3' was subcloned into pMON33510 creating pMON41417, pMON41418, and pMON41419, respectively. The details of the vector construction are shown in Figure 6. The antibody construct expressed in tobacco cell lines is also shown in Figure 6.

To analyze glycosylation patterns in monocot cells, a corn cell culture derived from an elite inbred line obtained from DEKALB (Mystic, CT) was used as a host system. Corn cell cultures were co-bombarded with the pMON41409 construct (the expression cassette for the antibody huNR-LU-10) and the pMON41411, pMON41412, or pMON41413 constructs, using a particle gun bombardment method. Control cells were bombarded solely with pMON41409. After bombardment, the transgenic calli were grown on 3.0 mM glyphosate selection media for 3 to 4 weeks. The calli were then screened by ELISA to assess the expression of the huNR-LU-10 antibody or by RNA blot analysis or a GalT enzyme assay to determine the expression of GalT. For each construct, approximately 25 independent transgenic calli were selected and screened for antibody expression by ELISA and Western blotting.

The tobacco cell line BY2 (*N. tabacum L. cv.* Bright Yellow 2) expressing huNR-LU-10 antibody was utilized to evaluate glycosylation patterns in dicot cells. Cells were grown in a 250 ml baffled flask for approximately 4 days and then used for bombardment. A thin layer of cells was placed onto a 150 μm mesh polypropylene screen (15 mm diameter) that had been positioned on a water agar plate. After the media was drained, the cell disc was osmotically stressed by incubation in 0.3 M osmoticum media for 1 hour. For the bombardment method, DNA for the huNR-LU-10 antibody was precipitated onto gold beads at the concentration of 1.0 μg DNA/mg gold using $CaCl_2$ and spermidine. The beads were washed and then resuspended in ethanol. Approximately 163 μl of the DNA/gold bead mixture was layered onto a mylar sheet. The DNA was blasted into the tobacco cells using a particle gun (14 kv, 14 vac). Following the DNA bombardment, the cells were incubated for approximately 24 hours in the dark at 25°C. The cell discs were then transferred to media containing 0.1 M osmoticum and incubated in the dark for an additional 24 hours at 25°C. The cell discs were then placed in tobacco suspension media containing 0.5 mM glyphosate media and 350 mg/L kanamycin. The cells were transferred to fresh selection media once per week until yellowish clumps of transformed callus (3 to 4 weeks) appeared. Individual

clones were allowed to proliferate 2 to 3 weeks on selection media and then screened by either RNA blot analysis or GalT activity assays. Approximately 30 independent transgenic calli for each construct were selected and screened for the expression of galactosyltransferase.

Positive calli were transferred to a 250 ml baffle flask containing 100 ml of tobacco cell suspension media (4.31 g/L MS Salts; thiamin; niacin; pyridoxine; glycine (250 μg/L); 87 mM sucrose; 0.5 mM glyphosate; 0.2 μg/ml 2,4-D; pH 5.8) or corn cell culture media (4.31 g/L MS Salts; 1 mg/L 2,4-D; thiamin; 1 M sodium EDTA; 200 mg/L casein hydrolysate; 0.25 mM proline; 2.8 mM potassium sulfate; 2.9 mM potassium phosphate monobasic; 0.5 mM myo-inositol; 87 mM sucrose; 3.0 mM glyphosate; pH 6.0). For the scale production, the cell suspension (90 ml) was mixed with 310 ml of media in a 1.0 L baffle flask (4 baffles: 3.5 cm wide x 1.0 cm deep). For the time course analysis, 30 ml of packed cell volume (pcv) was mixed with 60 ml of conditioned media and 310 ml of either corn cell culture media or tobacco cell suspension media in a 1.0 L flask. The flask cultures were sub-cultured once a week and grown at 28°C with a shaker speed of 125 rpm.

Total RNA from transgenic corn and tobacco calli was extracted using Trizol reagent (Life Technologies, Rockville, MD). Approximately 300 mg of plant tissue was ground with a mortar and pestle under liquid nitrogen. The fine powder was transferred to a fresh tube and 1 ml Trizol reagent was added. After vortexing for 10 seconds, the samples were incubated at room temperature for 5 minutes. Insoluble material was removed by centrifugation at 11,000 rpm for 10 minutes at 4°C. The supernatant was transferred to a fresh tube containing 200 μl of chloroform, vortexed for 15 seconds, and incubated at room temperature for 3 minutes. The samples were again centrifuged at 11,000 rpm for 10 minutes at 4°C. The upper aqueous phase (approximately 600 μl) was transferred to a fresh tube containing 500 μl isopropanol, incubated at room temperature for 10 minutes, and then centrifuged at 11,000 rpm for 1 minute at 4°C. The RNA pellet was washed with 70% ethanol and dissolved in 25 μl of DEPC water. The concentration of total RNA was measured (1 absorbance (A260) unit = 40 μg/ml RNA).

Corn and tobacco calli were analyzed for the presence of GalT RNA by RT-PCR and Northern blot. An example of the RT-PCR analysis is demonstrated in Figure 7. An example of the Northern screening of corn and tobacco is illustrated in Figures 8 and 9, respectively.

Under the stringency conditions of these experiments, the mouse probe from pMON41417 cross-hybridized to the bovine (pMON41418) sequence.

Approximately 100 mg of transgenic calli was homogenized using a handheld electric drill and then transferred to a microfuge tube containing 100 µl PBS. The samples were centrifuged in a tabletop microfuge at 4,000 rpm for 10 minutes. The supernatant was transferred to a 1.5 mL microfuge tube and analyzed for protein concentration using Coomassie assay (BioRad, Cambridge, MA).

Total soluble protein was extracted from corn transgenic calli and analyzed by ELISA to evaluate the expression level of the huNR-LU-10 antibody (Figure 10).

ELISA assays were performed according to standard procedures well known in the art. (Creighton, *Protein Structure: A Practical Approach*, Oxford University Press (New York, New York), 1989, pp. 99-101 and 288-310.)

Total soluble protein (20 µg) from transgenic calli or culture media (10 µl) was separated on a 4-20% SDS polyacrylamide gel (NOVEX Corporation, Carlsbad, CA) under either reduced or non-reduced conditions with a constant current of 40 mA. The samples were transferred to a Hybond C membrane (Amersham Pharmacia Biotech, Inc., Piscataway, NJ), and probed with anti-gamma antibody, followed by detection with anti-goat antibody conjugated to horseradish peroxidase (HRP) and chemiluminescence (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) for western blot analysis.

GalT activity assay was performed according to Schwientek et al., 271 J. BIOL. CHEM. 3398-405 (1996). Briefly, a 100 µl reaction consisting of 50 mM HEPES (pH 7.4); 20 mM $MnCl_2$; 54 mM NaCl; 0.5% NP40; 0.3 µCi UDP-[^{14}C] galactose (15Ci/mmol; American Radiolabeled Chemicals Inc., St. Louis, MO); 800 µg ovalbumin, and sample (10 µl) was incubated at 37°C for 2 hours. After incubation, the samples were precipitated with TCA, filtered, washed, and the amount of radiolabeled UDP was measured using a liquid scintillation counter. The enzyme activity was presented as units of galactose per µg of total soluble protein.

Approximately 500 mg of calli was homogenized with a mortar and pestle in 500 µl buffer M (25 mM Tris-HCl (pH 7.2); 0.5 mM DTE; 1 mM EDTA; 0.5 mM PMSF; and 250 mM sucrose), transferred to a microfuge tube, and further homogenized with a polytron for 30 seconds (10-second homogenization, 10 seconds on ice). The samples were transferred to

a 15 ml polypropylene tube and diluted to 10 ml with buffer M. Following filtration through cheese cloth, the samples were centrifuged at 3000x g for 15 minutes at 4°C to remove cell debris, nuclei, and mitochondria. The supernatant was centrifuged at 100,000x g for 1 hour at 4°C using an ultracentrifuge. The final pellet was dissolved in 100 µl buffer M and used for GalT activity analysis.

Total soluble protein or the Golgi membrane fraction was isolated from transgenic calli that demonstrated an accumulation of β(1,4)-galactosyltransferase (GalT) mRNA. The protein fractions were used to analyze GalT activity (Figures 11 and 12).

The transgenic calli were harvested, suspended in extraction buffer (25 mM Tris-HCl (pH 7.4); 0.25 M sucrose; 1 mM $MgCl_2$; 50 mM KCl) and then disrupted by sonication or homogenization. The protein concentration was measured by the Coomassie assay (BioRad) with BSA (Sigma, St. Louis, MO) as the standard.

GalT activity was assayed using UDP-Gal and pyridylamino-labeled $GlcNAc_2Man_3GlcAc_2$($GlcNAc_2Man_3GlcAc_2$-PA) as the substrate. The enzyme reaction contained 1-120 µg protein, 25 mM sodium cacodylate (pH 7.4), 10 mM $MnCl_2$, 200 uM UDP-Gal, and 100 nM $GlcNAc_2Man_3GlcAc_2$-PA. The reaction products were analyzed by HPLC using a PALPAK Type R and a Type N column (Takara Shuzo) according to the manufacturer's recommendations. The PA-labeled standards, $GlcNAc_2Man_3GlcAc_2$-PA, $Gal_2GlcNAc2Man_3GlcAc2$, and two isomers of $GalGlcNAc_2Man_3GlcAc_2$-PA, were obtained from Takara Shuzo and Honen Co. (Tokyo, Japan), respectively.

Purified antibodies, for example huNR-LU-10, isolated from the corn-GalT transgenics (mouse GalT, bovine GalT, and human GalT) and tobacco-GalT transgenics (mouse GalT, bovine GalT, and human GalT) are analyzed to determine whether galactose has been added to the glycan chain of the antibody.

The structures of the N-linked glycans of the huNR-LU-10 antibody from transformed cells are analyzed by a combination of reverse-phase and size-fractionation HPLC (RP- and SF-HPLC, respectively), two-dimensional sugar chain mapping, exoglycosidase digestions, and IS-MS/MS.

The glycosidase digestions with either β-N-acetlyglucosaminidase (*Diploccocus pneumoniae*; Boehringer Mannheim, Indianapolis, IN) or α-mannosidase (jackbean, Sigma, St. Louis, MO) are performed using 1 nmol of PA-sugar chains. For the β-galactosidase (*D.*

pneumoniae, Boegringer Mannheim, Indianapolis, IN) and α-(1,2)-mannosidase (*Aspergillus*) digestions, PA-sugar chains (1 nmol) in 50 mM sodium acetate buffer (pH 5.5) are incubated with either 200 mU β-N-acetlyglucosaminidase or 60 μg α-(1,2)-mannosidase at 37°C. The reactions are terminated by boiling, and an aliquot of the digests are analyzed by SF-HPLC. The molecular masses of the resulting digests are analyzed by IS-MS/MS and/or compared with control (animal-derived and/or unmodified plant-derived) sugar chains.

The IS-MS/MS experiments are performed using a Perkin-Elmer Sciex API-III, triple-quadrupole mass spectrometer (Perkin-Elmer, Norwalk, CT). The scanning is accomplished with a step size of 0.5 Da and the collisionally activated dissociation daughter ion spectrum is recorded from *m/z* 200.

Example 2. Expression of Immunoglobulins with Mammalian-like Glycosylation in Corn Plants

In another embodiment of the invention, bovine GalT was expressed in an endosperm-specific expression cassette for production in corn seed. Plasmid pMON41522 was created from pMON41402 using standard cloning techniques well known in the art. The sequence of bovine GalT has been reported previously by D'Agostaro et al., 183(1) EUR. J. BIOCHEM. 211-17 (1989) (GenBank accession number X14558). The pMON41522 plasmid places the expression of bovine GalT, generated by reverse transcriptase PCR of bovine liver polyA+ RNA as described above, under the control of the rice glutelin-1 promoter and leader. The Gt1 promoter/leader has been shown to promote high levels of heterologous polypeptide expression in maize endosperm. *See* WO 98/10,062. This embodiment of the invention may be advantageous because it is tissue-specific. Hence, the plant may not be adversely affected in a systemic fashion by the activity associated with a particular post-translational modification.

Other regulatory elements of pMON41522 include the HSP70 intron (WO 98/10062), the 5'UTR leader sequence of bovine GalT and its 3' UTR, followed by the RUBISCO SSU 3'UTR. The plasmid selection marker is Kanamycin. Thus, pMON41402 was cut with XbaI and BclI and the vector fragment isolated. Subsequently, the DNA encoding the gene for the *N. tabacum* GnT1 was isolated from pMON41530 as a BamHI-XbaI fragment. The expression vector pMON41522 was formed upon ligation of the two fragments. The expression cassette of pMON41522 is depicted in Figure 13.

The plasmid was then introduced into corn cell material by bombardment as previously described in Example 1. *See also* WO 98/10,062. Prior to use, the plasmid vector was cut with the appropriate restriction enzymes at sites on either side of the plant transgene cassette. The expression cassette is then purified to eliminate bacterial sequences. The corn cell material used is that with was competent to regenerate fertile corn plants. Gordon-Kamm et al., 2 PLANT CELL 603-18 (1990). After transformation, the corn material was placed on kanamycin selection media. The surviving cells were placed into a series of media conditions of varied salts and plant growth regulators, to stimulate the organizated production of plant roots and shoots. *See, e.g.,* Green & Phillips, 14 CROP. SCI. 417-20 (1975); WO 98/10,062. The plantlets are then planted in the appropriate soil mixture. Plants are grown to maturity in the greenhouse, self pollinated, and the seed harvested. Seed can be replanted and grown. Plants can be re-selfed, so that a pure-breeding transgenic trait is developed, or outcrossed, to put the transgene in a novel genetic background, or create more transgenic material by transferring the transgenic pollen to more nontransgenic ears.

Mature seed are pulverized individually or as a pool, then extracted in aqueous buffer and the solids removed by centrifugation. Total protein is determined by Coomassie dye binding assay (Bio-Rad Laboratories, Hercules, CA) or BCA assay (Pierce Chemical Co., Rockford, IL) with bovine IgG as a standard. Extracts are screened by ELISA and western methods as above.

Alternatively, protein is extracted from dry seed and endosperm as follows: Surface sterilized seeds are placed on wet sterile filter paper overnight at room temperature to imbibe water. The scutellum is then removed, and the part of the kernel remaining after removing the scutellum is treated as endosperm because the remaining pericarp contains only 4% of total soluble protein and 2% of the total antibody.

Total soluble protein from the endosperm is essentially extracted the same way as the whole seed except the endosperm is ground with pestle and mortar.

The following extraction buffer is useful for the extraction of protein from dry seeds and endosperm: for 500 ml of buffer, add 100 ml 0.5 M $NaPO_4$ (pH 7.0), 20 ml 0.5 M EDTA (pH 8.0), 10 ml 10% Triton X-100, and 370 ml dH_20 (final concentration: 100 mM $NaPO_4$, 20 mM EDTA, 0.2% Triton X-100). The buffer should be filter-sterilized or made from sterile stocks.

Individual maize seeds are crushed between layers of glassine weighing papers on an Arbor seed press. Ground material is poured into a 1.5 ml microcentrifuge tube and approximately 1.5 ml extraction buffer is added to the tube. The mixture is vortexed on high for 5-10 seconds and incubated at 4°C for at least 2 hours. The extractions are then sonicated for 1.5 to 2 minutes. The tubes are centrifuged in a microcentrifuge at the highest speed for 20 minutes at 4°C to remove cell debris. The supernatant is transferred to a fresh tube. A set of protein standards is prepared using 1 μg/μl human IgG. The amount of protein in the standards and the unknowns is measured using the Coomassie Plus Protein Assay kit (Pierce Chemical Co., Rockford, IL). Absorbance is read at 595 nm. Unknowns are compared to an IgG standard curve to determine total protein in the seed extract.

GalT activity assay is performed according to Schwientek et al., 271 J. BIOL. CHEM. 3398-405 (1996) as described in detail in Example 1. The amount of radiolabeled UDP is measured using a liquid scintillation counter. The enzyme activity is presented as units of heterologous polypeptide per μg of total soluble protein.

GalT-expressing corn plants are cross-pollinated with corn plants that express a cloned human monoclonal antibody. This antibody is comprised of both the heavy and light immunoglobulin chains, such as hu-NR-LU-10 as described in Example 1. Corn expressing single-chain antibody may be used in an alternative approach. After pollination, seed are allowed to ripen and are then harvested. Several seeds from a given ear of corn are analyzed, by standard techniques described below, for altered glycosylation of the antibody molecules produced therein.

Purified antibodies produced in the corn-GalT transgenics (for example huNR-LU-10) are analyzed to determine whether galactosyl residues are added to the glycan chain of the antibody. Other glycosylated target heterologous polypeptides are analyzed in similar fashion.

The structures of the N-linked glycans of the huNR-LU-10 antibody from transformed cells are analyzed by a combination of reverse-phase and size-fractionation HPLC (RP- and SF-HPLC, respectively), two-dimensional sugar chain mapping, exoglycosidase digestions, and IS-MS/MS.

The glycosidase digestions with either β-N-acetlyglucosaminidase (*Diploccocus pneumoniae*; Boehringer Mannheim, Indianapolis, IN) or α-mannosidase (jackbean, Sigma, St. Louis, MO) are performed using 1 nmol of PA-sugar chains. For the β-galactosidase (*D.*

pneumoniae, Boegringer Mannheim, Indianapolis, IN) and α-(1,2)-mannosidase (*Aspergillus*) digestions, PA-sugar chains (1 nmol) in 50 mM sodium acetate buffer (pH 5.5) are incubated with either 200 mU β-N-acetylglucosaminidase or 60 μg α-(1,2)-mannosidase at 37°C. The reactions are terminated by boiling, and an aliquot of the digests are analyzed by SF-HPLC. The molecular masses of the resulting digests are analyzed by IS-MS/MS and/or compared with control (animal-derived and/or unmodified plant-derived) sugar chains.

The IS-MS/MS experiments are performed using a Perkin-Elmer Sciex API-III, triple-quadrupole mass spectrometer (Perkin-Elmer, Norwalk, CT). The scanning is accomplished with a step size of 0.5 Da and the collisionally activated dissociation daughter ion spectrum is recorded from *m/z* 200.

Example 3: GalT Enzyme for Altering the Glycosylation Pattern of Immunoglobulin Molecules Produced in Plants

As noted previously, an approach to effecting post-translational modification involves targeting a heterologous polypeptide to a particular organelle in order to effect its modification. Additionally, a post-translational modification enzyme may be introduced, modified or non-modified, to direct its expression and function to a desired location within an organelle, which will have bearing on both the spatial and temporal interactions of both it and various other modifying enzymes. Thus, for example, a heterologous β-1,4-galactosyltransferase (GalT) (EC 2.4.1.38) can be introduced into the plant cell for interactions in the Golgi with a heterologous polypeptide, such as an antibody, that will consequently have a more mammalian-like glycosylation pattern. *Compare* Figures 1, 2 and 16.

More specifically, the sequential reaction of glycan trimming/biosynthesis (Figure 16) in the lumen of the ER and in the Golgi compartments is a consequence of the spatial distribution of the enzymes in the corresponding organellar membranes. That distribution, on the other hand, is determined by the nature of the cytosolic, membrane-spanning, and stem regions (CMS regions) of the glycosyltransferase enzymes that precede, N-terminally, the active catalytic domain. The CMS region determines not only the spatial positioning of the enzymes, but perhaps also the half-life of the membrane-bound active form. *See* ESSENTIALS IN GLYCOBIOLOGY, (Varki et al., eds., 1999). Because plants do not possess GalT, the mammalian CMS region may or may not function in an optimized way to position bovine

GalT within the sequential path of plant glycosyltransferase and hydrolase enzymes to exert galactose addition to a heterologous protein. Plant CMS regions from other class-II membrane-spanning glycosyltransferases and hydrolases may not be more suited to efficiently mediate the positioning of the catalytic domain of GalT along the intracellular path.

This embodiment of the invention, then, replaces the CMS region of the mammalian GalT, to be transformed into plants, with the CMS region of an earlier enzyme in the pathway, such as GnT1 or mannosidase 1. This exchange positions the GalT enzyme alongside the latter enzymes that trim the glycan residues for subsequent fucosyl and xylosyl addition. In other words, the modified GalT competes with and is active along side the enzymes that prepare the glycan structure for the fucosyl and xylosyl residues. As a result, immediate action of the modified GalT may sterically or spatially prevent the addition of fucosyl and xylosyl residues. This effect is further enhanced if the heterologous protein recycled between the ER and the pre-Golgi (and possibly cis-Golgi) compartment, with the latter demonstrated to add N-acetylglucosamine (GlcNAc), a precondition for galactose addition. Gomord et al., 81 BIOCIMIE 607-18 (1999). The retention in the pre-Golgi (and possibly cis-Golgi) compartment shall prevent the addition of α-1,2-xylose in the medium-Golgi compartment, and of β-1,3-fucose in the trans-Golgi compartment, and yields a heterologous polypeptide, such as an antibody, with a more mammalian-like glycosylation pattern. Such antibodies may be less immunogenic in humans than native plant-glycosylated polypeptides.

While the amino acid sequence of the catalytic domains of GalT remain relatively constant between plant species, the amino acids comprising the CMS regions do not. This is easily seen by aligning the published sequences for this gene. Identifying the variable CMS-encoding region indicates which portion of the GalT gene may be substituted with the CMS region from an early-pathway plant glycosyltransferase.

This approach is useful in both monocots and dicots, whether in cell culture or plants, for any target heterologous polypeptide for which altered post-translational modification is desired.

Procedures

GalT was modified such that its N-terminus was replaced with the N-terminus of tobacco GnT1. More specifically, a plasmid carrying a hybrid GnT1/GalT, pMON41523,

was constructed. The 5' regulatory elements of this construct include the endosperm-specific rice glutelin 1 promoter and leader (Gt1), followed by HSP70 and the signal/leader sequence of the extensin 5' UTR. The DNA encoding the 79 N-terminal amino acids of *N. tabacum* GnT1 (GenBank accession number Y16832), was placed upstream of the bovine GalT gene (GenBank accession number X14558) from which a 75-amino acid-encoding XbaI-SacI fragment had been removed. Thus, pMON41522 was cut with XbaI and SacI and the vector fragment isolated. Subsequently, the DNA encoding the N-terminal 79 amino acids of the gene for the *N. tabacum* GnT1 was isolated from pMON41530 as a XbaI-SacI fragment. Ligation of these two fragments resulted in the expression cassette of pMON41523 (Figure 15).

Another bovine GalT was modified by replacing the N-terminal 75 amino acids with 38 amino acids from the N-terminus of *A. thaliana* α-1,2-mannosidase (EC 3.2.1.113). Specifically, the plasmid pMON41587, carrying the hybrid gene, was created by standard techniques described above for pMON41523. The regulatory regions in this construct are analogous to pMON41523, above (Figure 15). The DNA encoding the N-terminal 39 amino acids of the gene for the *A. thaliana* α-1,2-mannosidase (GenBank accession number AC005916) was isolated as a XbaI-SacI fragment. This fragment was ligated to the bovine GalT gene (X14558) from which DNA encoding the 75 N-terminal amino acids had been cleaved via XbaI-SacI.

An additional bovine GalT was modified by replacing the N-terminal 75 amino acids with 106 amino acids from the N-terminus of *Glycine max* α-1,2-mannosidase (EC 3.2.1.113). Specifically, the plasmid pMON41588, carrying the hybrid gene, was created by standard techniques described above for pMON41523 and pMON41587. The regulatory regions are analogous to pMON41523 and pMON41587, above (Figure 15). The N-terminal 106 amino acids of the gene encoding soybean α-1,2-mannosidase (GenBank accession number AC005916) was isolated as an XbaI-SacI fragment, and ligated to the bovine GalT (X14558) gene from which 75 N-terminal amino acids had been cleaved via XbaI-SacI.

GalT modifications are illustrated in multiple alignments as shown in Figure 14. where MNS285 represents the bovine GalT, GNT_GalT represents the hybrid molecule of plasmid pMON41523, MAN_ATGT represents plasmid pMON41587, and MAN_GMGT represents the hybrid molecule of plasmid pMON41588.

In another example, instead of constructing an expression vector using the endosperm-specific rice glutelin 1 promoter for stable expression in plants, the 35S constitutive promoter may be used to study plant cells or tissues on a short term or more rapid basis. See Example 1.

As described in Example 2, corn cell material is transformed with one of the pMON41523, pMON41587 or pMON41588 expression constructs, using a particle gun bombardment method. Prior to use, the plasmid vector is cut with the appropriate restriction enzymes at sites on either side of the plant transgene cassette. The expression cassette is then purified to eliminate bacterial sequences. The corn cell material used is that with is competent to regenerate fertile corn plants. After transformation, the corn material is placed on kanamycin selection media. The surviving cells are placed into a series of media conditions of varied salts and plant growth regulators, to stimulate the organizated production of plant roots and shoots. The plantlets are then planted in the appropriate soil mixture. Plants are grown to maturity in the greenhouse, self pollinated, and the seed harvested.

These corn cell materials may be co-bombarded with the pMON41409 construct (the expression cassette for the antibody huNR-LU-10) or another target heterologous protein expression construct. Alternatively, the GalT expression cassettes may be introduced into tissues that already express the target protein. As another alternative approach, plants expressing the hybrid GalT construct may be sexually crossed with a plant expressing the target heterologous polypeptide.

Total RNA from transgenic corn samples are extracted using detailed procedures described in Example 1. Powdered calli is extracted using Trizol reagent. After vortexing, incubation and centrifugation, the resulting supernatant is extracted with chloroform. The upper aqueous phase of this second extraction is transferred to a fresh tube containing isopropanol, incubated and then centrifuged. The resulting RNA pellet is washed with 70% ethanol and dissolved in DEPC water. The concentration of total RNA is measured (1 absorbance (A260) unit = 40 μg/ml RNA). *Nicotiana tabacum* (NT) calli are analyzed for the presence of GnT1/GalT or Mann/GalT RNA by RT-PCR and Northern blot.

Leaf extracts from transgenic plants are screened for the presence of heterologous protein using an ELISA assay based on the methods described by Engvall et al., 109 J. IMMUNOL. 129-35 (1972). After removing the midvein, the leaves are homogenized in a mortar and pestle and then 0.25 volume 5x TBS (750 mM NaCl, 100 mM Tris-HCl , pH 8.0)

is added to the homogenized leaves. Two-fold serial dilutions of the homogenate are prepared in TBS (150 mM NaCl, 20 mM Tris-HCl , pH 8.0) and aliquots (50 µl) of the serial dilutions are then used for protein analysis.

For Western blots, leaf segments (1 gram) from mature plants are homogenized in a mortar and pestle with 1 ml homogenization buffer (0.05 M Tris-HCl (pH 7.5), 1 mM phenylmethylsulfonyl fluoride). Following homogenization, the resulting leaf extract is suspended in buffer containing 4 M urea and 1% SDS with or without 2 mM DTT depending on the reducing conditions. The solution is boiled for 3 minutes. After boiling, the solution is loaded onto a 10% polyacrylamide gel for electrophoresis (SDS-PAGE) and then the electrophoresed proteins are transferred to nitrocellulose for protein detection.

Alternatively, mature seed are pulverized individually or as a pool, then extracted in aqueous buffer and the solids removed by centrifugation. Total protein is determined by Coomassie dye binding assay (Bio-Rad Laboratories, Hercules, CA) or BCA assay (Pierce Chemical Co., Rockford, IL) with bovine IgG as a standard. Extracts are screened by ELISA and western methods as above.

Alternatively, protein is extracted from dry seed and endosperm as follows: Surface sterilized seeds are placed on wet sterile filter paper overnight at room temperature to imbibe water. The scutellum is then removed, and the part of the kernel remaining after removing the scutellum is treated as endosperm because the remaining pericarp contains only 4% of total soluble protein and 2% of the total antibody.

Total soluble protein from the endosperm is essentially extracted the same way as the whole seed except the endosperm is ground with pestle and mortar.

The following extraction buffer is useful for the extraction of protein from dry seeds and endosperm: for 500 ml of buffer, add 100 ml 0.5 M $NaPO_4$ (pH 7.0), 20 ml 0.5 M EDTA (pH 8.0), 10 ml 10% Triton X-100, and 370 ml dH_20 (final concentration: 100 mM $NaPO_4$, 20 mM EDTA, 0.2% Triton X-100). The buffer should be filter-sterilized or made from sterile stocks.

Individual maize seeds are crushed between layers of glassine weighing papers on an Arbor seed press. Ground material is poured into a 1.5 ml microcentrifuge tube and approximately 1.5 ml extraction buffer is added to the tube. The mixture is vortexed on high for 5-10 seconds and incubated at 4°C for at least 2 hours. The extractions are then sonicated for 1.5 to 2 minutes. The tubes are centrifuged in a microcentrifuge at the highest speed for

20 minutes at 4°C to remove cell debris. The supernatant is transferred to a fresh tube. A set of protein standards is prepared using 1 μg/μl human IgG. The amount of protein in the standards and the unknowns is measured using the Coomassie Plus Protein Assay kit (Pierce Chemical Co., Rockford, IL). Absorbance is read at 595 nm. Unknowns are compared to an IgG standard curve to determine total protein in the seed extract.

GnT1/GalT or Mann/GalT activity assay is performed according to Schwientek et al., 271 J. BIOL. CHEM. 3398-405 (1996) as described in detail in Example 1. The amount of radiolabeled UDP is measured using a liquid scintillation counter. The enzyme activity is presented as units of heterologous polypeptide per μg of total soluble protein.

Hybrid GalT-expressing corn plants are cross-pollinated with corn plants that express a cloned human monoclonal antibody. This antibody is comprised of both the heavy and light immunoglobulin chains, such as hu-NR-LU-10 as described in Example 1. Corn expressing single-chain antibody may be used in an alternative approach. After pollination, seed are allowed to ripen and are then harvested. Several seeds from a given ear of corn are analyzed, by standard techniques described below, for altered glycosylation of the antibody molecules produced therein.

Purified antibodies produced in the corn-GalT transgenics (for example huNR-LU-10) are analyzed to determine whether galactosyl residues are added to the glycan chain of the antibody. Other glycosylated target heterologous polypeptides are analyzed in similar fashion.

The structures of the N-linked glycans of the huNR-LU-10 antibody from transformed cells are analyzed by a combination of reverse-phase and size-fractionation HPLC (RP- and SF-HPLC, respectively), two-dimensional sugar chain mapping, exoglycosidase digestions, and IS-MS/MS.

The glycosidase digestions with either β-N-acetylglucosaminidase (*Diploccocus pneumoniae*; Boehringer Mannheim, Indianapolis, IN) or α-mannosidase (jackbean, Sigma, St. Louis, MO) are performed using 1 nmol of PA-sugar chains. For the β-galactosidase (*D. pneumoniae*, Boegringer Mannheim, Indianapolis, IN) and α-(1,2)-mannosidase (*Aspergillus*) digestions, PA-sugar chains (1 nmol) in 50 mM sodium acetate buffer (pH 5.5) are incubated with either 200 mU β-N-acetylglucosaminidase or 60 μg α-(1,2)-mannosidase at 37°C. The reactions are terminated by boiling, and an aliquot of the digests are analyzed by SF-HPLC.

The molecular masses of the resulting digests are analyzed by IS-MS/MS and/or compared with control (animal-derived and/or unmodified plant-derived) sugar chains.

The IS-MS/MS experiments are performed using a Perkin-Elmer Sciex API-III, triple-quadrupole mass spectrometer (Perkin-Elmer, Norwalk, CT). The scanning is accomplished with a step size of 0.5 Da and the collisionally activated dissociation daughter ion spectrum is recorded from *m/z* 200.

Example 4. Antisense Molecules Targeting Glycotransferases

As described in greater detail previously, plant-specific glycosylation and complex-type N-glycan formation is not always desired for a given heterologous polypeptide. Genes and mRNAs encoding glycosyl transferases may be used to generate antisense nucleic acids that epigenetically interfere with the production of these enzymes and thus reduce the activity of native enzymes and alter the natural post-translational modification abilities of the host system. Another approach includes the generation of stable plant production hosts through other mechanisms of epigenetic silencing such as combined antisense-sense constructs, or the generation of mutants either chemically (i.e. EMS) or through transposable genetic elements.

More specifically, the pathway to plant-specific glycosylation and complex-type N-glycan formation of glycoproteins is schematically depicted in Figure 16. Briefly, after transfer of the basic glycan structure, GlcNAc3Man9Gluc3, by oligosaccharyltransferase (OST), from dolicholpyrophosphate to the Asn-X-Ser/Thr recognition sequence of the nascent polypeptide chain of a secretory protein, glucosidases and mannosidases in the lumen of the endoplasmic reticulum start to trim back the extent of the glycans. In the Golgi, GnT1 adds N-acetylglucosamine to the peripheral mannose residue; and not before this reaction is completed an α-1,3 fucosyl residue is attached to the proximal GlcNAc residue of the core. Further mannose trimming and GlcNAc addition is followed by β-1,2-xylose addition, forming a potentially immunogenic complex-type N-glycan structure.

The importance in the plant pathway of the enzyme N-acetylglucosamine transferase (GnT1) has been shown. Specifically, an *Arabidopsis* mutant *cgl1*, that lacked the presence of complex glycans, was studied by Von Schaewen et al, 102 PLANT PHYSIOL. 1109-18 (1993). The ability of this mutant to synthesize complex glycan structures was restored by transforming the plant with a human GnT1 gene, thus confirming the key triggering role of GnT1 in complex glycan biosynthesis in plants. Gomez & Chrispeels, 91 P.N.A.S. 1829-33

(1994). Hence, antisense molecules targeted against enzymes in this pathway, such as GnT1, should effect profound changes in plant glycosylation patterns.

For example, GnT1 (N-acetylglucosamine transferase 1: E.C.2.4.1.101), is cloned and used for an antisense molecule. The sequence of tobacco, potato, and *Arabidopsis* GnT1 genes have been reported. WO 99/29879; Strasser et al., 9 GLYCOBIOLOGY 779-85 (1999). This gene has also been sequenced for corn and soybean. Antisense constructs developed with the desired regulatory elements are introduced (for example, by transformation or sexual crossing) into a plant that already expresses the heterologous glycoprotein of interest. Screening of knock-out mutants (by Northern Blotting, PCR, and Western techniques) is used to screen transformants encoding the glycoprotein of interest. This approach quickly yields evidence of proof of concept.

As another example of this approach, diminishing the activity of enzymes further down the glycan maturation path also leads to the prevention of complex-type N-glycan formation of glycoproteins in plants. For example, the α-1,3-linked fucosyl residue to GlcNac is the predominant antibody determinant in plant glycoproteins. Together with β-1,2-linked xylosyl residues, these complex-type glycans are also called "cross-reactive carbohydrate determinants." A core fucosyltransferase from mung bean has been reported by Leiter et al., 274 J. BIOL. CHEM. 21830-39 (1999). Homologous clones have been identified in corn and soybean databases. Additionally, animal sequences for this enzyme are known, including human, rat, and *C. elegans*. The biochemical characterization of another downstream enzyme, soybean beta-1,2-xylosyltransferase (XylT), along with a partial DNA sequence, has been reported. WO 99/29835. Similarly, a homologous clone of XylT has been reported in *Arabidopsis*. Strasser et al., 452 FEBS LET. 105-08 (2000). The sequence of XylT has also been identified in maize.

Whether the target enzyme appears early or later in the glycan biosynthesis pathway, the approach to test a particular target is standard. After the determination of the native copy number of the gene, the antisense construct may give fast proof of concept and enable the technology, followed by knock-out mutant search and/or epigenetic silencing approaches targeting these glycosyltransferases.

The methods of constructing an antisense vector are standard. Known sequences of analogous enzymes are identified, and if they are not ideal for expression, then target

sequences are identified from appropriate databases. Clones are verified, the antisense sequence defined, and then the appropriate sequences placed into an expression vector.

Procedures

An antisense construct targeting maize GnT1 was constructed. The GnT1 sequence was identified by matching the published tobacco sequence (GenBank accession number Y168332) with homologous sequences in a *Zea mays* gene bank. This technique identified a sequence of about 144 amino acids, of maize, that corresponds to the C-terminal region of the tobacco gene which encodes about 445 amino acids (GnT1-As). In the resulting plasmid, pMON41581, the antisense transcript was placed under the regulatory control of the endosperm-specific rice glutelin 1 promoter and leader, and the HSP70 intron. The 3'end of the transcript was polyadenylated via a soybean RUBISCO SSU 3'UTR (Figure 17). Thus, pMON41230 was cut with BamHI and XbaI and the vector fragment isolated. Subsequently, the DNA encoding the gene for the antisense gene was isolated as a XbaI-BglII fragment. The expression vector pMON41581 was formed upon ligation of the two fragments.

An antisense construct targeting beta-1,2-xylosyltransferase (XylT) was prepared by cloning the antisense *A. thaliana* (XylT) gene (GenBank accession number AJ272121) under control of the rice glutelin 1 promoter and leader. The expression of the XylT antisense in this plasmid, pMON41515, is also regulated upstream by HSP70, and downstream by a GUS insert (Figure 17). Thus, pMON41230 was cut with BamHI and XbaI and the vector fragment isolated. Subsequently, the DNA encoding the gene for the antisense gene was isolated as a XbaI-BamHI fragment. The expression vector pMON41515 was formed upon ligation of the two fragments.

An antisense construct targeting fucosyl transferase (FucT) was constructed using the homologous antisense sequence of *Vigna radiaia* (GenBank accession number Y18529). This plasmid, pMON41582 (Figure 17), employs the same regulatory elements as pMON41515. Thus, pMON41230 was cut with BamHI and XbaI and the vector fragment isolated. Subsequently, the DNA encoding the gene for the antisense gene was isolated as a XbaI-BamHI fragment. The expression vector pMON41582 was formed upon ligation of the two fragments. *Z. mays* FucT, which has a high level of homology with the *V. radiata* sequence, is also used as an antisense construct in this same manner.

Other constructs may be prepared that do not employ the endosperm-specific rice glutelin 1 promoter. For example, the CaMV 35S promoter may be used to provide expression in plant cell or tissue cultures. See Examples 1 and 6.

The plasmid constructs (pMON41581, pMON41515 or pMON41582) are then introduced into corn callus-like tissue by bombardment as was described in Example 1. The corn materials are bombarded with one of the constructs (pMON41581, pMON41515 or pMON41582), using a particle gun bombardment method. As a null control, a callus blast is also done using a blasting sheet containing no DNA. Independent transgenic calli are identified by survival on selection media.

Host cells may also be co-transformed with a vector that expresses the target heterologous polypeptide, such as an antibody. Alternatively, corn materials are regenerated into mature plants and then crossed with plants that express the target heterologous polypeptide generating plants that express both genes.

Positive calli are transferred to a 250 ml baffle flask containing 100 ml of corn cell culture media. For the scale production, the cell suspension (90 ml) is mixed with 310 ml of media in a 1.0 L baffle flask (4 baffles: 3.5 cm wide x 1.0 cm deep). For the time course analysis, 30 ml of packed cell volume (pcv) is mixed with 60 ml of conditioned media and 310 ml of either corn cell culture media in a 1.0 L flask. The flask cultures are sub-cultured once a week and grown at 28°C with a shaker speed of 125 rpm.

SDS/PAGE and Protein Blotting: Total proteins are extracted by harvesting transgenic calli, suspending in extraction buffer (25 mM Tris-HCl (pH 7.4); 0.25 M sucrose; 1 mM $MgCl_2$; 50 mM KCl; and 1 mM PMSF), disrupting by either sonication or homogenization, and subsequently, centrifuging (Beckman-Spinco GS-15R) at 14,000 rpm for 10 min. Soluble proteins in the supernatant are quantified by Bio-Rad assay (Bio-Rad Laboratories, Hercules, CA). Proteins are separated on SDS/15% PAGE gels by using the Mini-Protean II system (Bio-Rad Laboratories, Hercules, CA), and the gels are stained with Coomassie brilliant blue R250 and examined for protein quality and quantity.

For Western blotting, the SDS/PAGE-separated proteins are transferred onto a nitrocellulose membrane (Schleicher & Schuell, Inc., Keene, NH). The blots are then probed with antibodies as described in CREIGHTON, PROTEIN STRUCTURE: A PRACTICAL APPROACH 105 (Oxford Univ. Press, New York 1989). Polyclonal antibodies are diluted using 0.5% BSA in PBS, and the binding of polyclonal antibodies are detected with anti-rabbit Ig-

conjugated alkaline phosphatase (Promega, Madison, WI). ELISA assays are performed according to standard procedures well known in the art. CREIGHTON, *supra* at 99-101, 288-310.

Total RNA is extracted from transgenic and wild type control corn samples by using the SNAP kit (Invitrogen Corporation, Carlsbad, CA) following the manufacturer's instructions. RNA preparations are quantified with a GeneQuant DNA/RNA calculator (Amersham Pharmacia Biotech, Inc., Piscataway, NJ). For RNA Northern blot analysis, total RNA (5μg) is electorphoresed and transferred onto Hybond N^+ nylon membrane (Amersham Pharmacia Biotech, Inc.). GnT1, XylT, and FucT sense and antisense is detected with ^{32}P labeled single-stranded RNA probes generated by *in vitro* transcription from pMON41581 (GnT1), pMON41515 (XylT), and pMON41582 (FucT) cDNA clones, respectively.

Transgenic plants are analyzed for expression of GnT1, XylT, and FucT by probing protein blots with GnT1, XylT, and FucT-specific polyclonal antibodies. Total protein extracted from transgenic and wild-type samples are analyzed by SDS/PAGE. No difference is observed in the overall soluble protein profile of transgenic and wild-type control plants. Use of GnT1, XylT, and FucT-specific polyclonal antibodies shows the presence of GnT1, XylT, and FucT protein in extracts from wild-type samples. In contrast, immunoreactive protein is reduced or absent in protein extracts of the transgenic samples. These results indicate specific inhibition of GnT1, XylT, and FucT production in transgenic plants.

Furthermore, Northern blot analysis of transgenic sample tissues indicates the presence of the expected size of the antisense transcripts, indicating that the repression of GnT1, XylT, and FucT is caused by the expression of the introduced antisense GnT1, XylT, FucT genes, respectively.

The approach thus described for corn is also appropriate for dicots such as soybean. In this approach, the soybean antisense for the target enzyme is identified from databases as described herein. The corresponding antisense DNA construct is placed under control of a species-appropriate promoter or a constutive promoter such as 35S CaMV promoter. McCormick et al., 5 PLANT CELL REPORTER. 81-84 (1986).

Example 5: Expression of a Biologically Active Post-Translationally Altered Human Protein C in Plant Cells

The γ-carboxylation of glutamic acid is another example of a mammalian post-translational modification that would be advantageously introduced to plant expression systems. This post-translational modification of glutamic acid to γ-carboxyglutamic acid, accomplished by γ-glutamyl carboxylase, is essential for the activity of vitamin K-dependent proteins such as several blood coagulation and anti-coagulation proteins as well as bone Gla (γ-carboxyglutamic acid) protein and bone matrix protein.

Human protein C (HPC) is a complex plasma glycoprotein that functions as an anticoagulant by proteolytically inactivating coagulation factors V_a and $VIII_a$. The protein C anticoagulant pathway is triggered when thrombin binds to the endothelial cell receptor, thrombomodulin. This complex activates protein C to generate the anticoagulant enzyme, activated protein C (APC), which in complex with protein S, inhibits coagulation by inactivating the critical regulatory proteins, factors V_a and $VIII_a$.

Fully functional protein C is produced by several post-translational modifications. Two enzymes required for this post-translational modification are the subtilisin-like serine protease furin, also known as paired basic amino acid cleaving enzyme (PACE), and γ-glutamyl carboxylase.

The protein C precursor is proteolytically modified to produce its mature form by PACE. Specifically, endoproteolytic processing of the protein C precursor involves cleavage of the propeptide after amino acids Lys^{-2}-$Arg4^{-1}$ and removal of a Lys^{156}-Arg^{157} dipeptide, yielding the light and heavy chains of the mature protein C dimer.

In addition, the vitamin K-dependent γ-carboxyglutamic acid (Gla) domain of activated protein C binds to the endothelium cell protein C receptor (EPCR). Specifically, the γ-carboxyglutamate residues of protein C are carboxylated by γ-glutamyl carboxylase to form binding sites with high affinity for calcium (Ca^{2+}). The binding of Ca^{2+} by functional protein C allows it to readily bind to the EPCR.

Although neither γ-glutamyl carboxylase nor PACE are present in plant cells, the necessary cofactor for γ-glutamyl carboxylase, vitamin K, is present in plant cells. Thus, plant vectors expressing these enzymes may be constructed to provide plant cells with these enzymatic functions.

Procedures

The DNA for human PACE and human γ-glutamyl carboxylase was obtained from ATCC (GenBank accession number 79823 and GenBank accession number 68666, respectively). The DNA for human protein C (HPC) was obtained from NRRL (GenBank accession number B-15926).

The PACE DNA was cloned into a plant expression cassette as follows: The initial ATCC clone was digested with EcoRI and EcoRV to release a 3 kb fragment (5' UTR segment, signal peptide, coding region, and 3' UTR segment) that was then ligated into the EcoRI to HincII site of the pUC19 cloning vector, to create pwrg4576. The XmnI to PstI fragment of pwrg4576 was generated to encompass the majority of the PACE mature coding region (truncated at the 5' end), and a smaller segment of the 3'UTR. This was ligated into the plant expression vector pwrg5087, that was digested with NcoI and PstI. Two linkers, dar19 (catggatgctcagggccagaagg) and dar21 (ccttctggccctgagcatc), were used to complete the 5' end of PACE. The final vector, pwrg4579, included the P-35S promoter, the tobacco extensin 5'UTR and signal peptide (with processing site), a C-terminally truncated PACE mature coding sequence, and the plant nopaline synthase (*nos*) 3' end.

To create the full length PACE in a plant expression vector, pwrg4579 was cut with PstI, blunt-ended with T4 DNA polymerase (Promega, Madison WI), and then cut with BglII within the PACE coding region. The PACE fragment from pwrg4576 from BglII to HinDIII was ligated into this vector, to create pwrg4631.

To create soluble PACE in a plant expression vector, a PCR reaction was performed using pwrg4631 and the primers dar56 (catgaattcccagaaagcagtggctgcaag) and dar57 (cgtcctaggtcatcactcaggcaggtgtgagggcag). After cutting with BsgI and AvrII, a C-terminal fragment was generated. This fragment contained a truncated coding region at the stop codons (*See* underlined regions of dar57, *supra*), which eliminated the transmembrane domain Rehemtulla et al., 79 BLOOD 2349-55 (1992). The truncated C-terminal fragment was then ligated into pwrg4631, to generate pwrg4643.

To create full-length PACE with the native signal peptide, vector pwrg4631 was cut with HinDIII, blunt-ended with T4 DNA polymerase (Promega, Madison, WI), and then digested with SacII. A fragment of pwrg4576 was created from SmaI and SacII digestion, to encompass the PACE 5' UTR, signal peptide, and a segment of the mature coding region of

PACE, which was then ligated into digested pwrg4631. The ligated product was called pwrg4644.

To create truncated, soluble PACE with the native signal peptide, vector pwrg4643 was cut with HinDIII, blunt-ended with T4 DNA polymerase, and then cut with SacII. A fragment of pwrg4576 was created from SmaI and SacII digestion, to encompass the PACE 5' UTR, signal peptide, and a segment of the mature coding region of PACE, which was then ligated into digested pwrg4643. The ligated product was called pwrg4645.

The human γ-glutamyl carboxylase vector pwrg4068 was cut at the Not I site downstream from the poly A site and a plant NPT2 selection cassette inserted, to create pwrg4604. The vector pwrg4069 was also created, which contains both the γ-glutamyl carboxylase cassette, and the GUS (β-glucuronidase) screenable expression cassette.

The plasmid, pHC7, containing the coding region for HPC (GenBank accession B-15926) was obtained from NRRL. An N-terminally truncated fragment of pHC7 was obtained by partial PstI digest, followed by digestion with BglII. Plasmid pHC7 was also used as a PCR template to generate a new fragment comprising the N-terminus of HPC. The PCR primers dar28 (tccgccatggcagctcctcttgac) and dar29 (gcaagaccaagcactggtcacc) were used to generate a N-terminus fragment of HPC and create a NcoI site. The PCR product was then cut with NcoI and BglII.

The N-terminal region of HPC containing the NcoI site and the N-terminally truncated fragment of HPC were ligated into the NcoI/PstI digested plant expression vector, pwrg5087, to create pwrg4590. The pwrg4590 vector now contained the P-35S promoter, the extensin signal peptide, HPC, and the *nos* poly A region.

To create HPC with its native signal peptide in a plant expression vector, pHC7 was cut with BanI and partially digested with PstI, which created a fragment with the entire HPC coding region. This fragment was ligated into the plant expression vector, pwrg5087, which was digested with HinDIII and PstI. A linker based on the AMV leader sequence was used to join the HinDIII site to the BanI site.

The final ligation product, pwrg4590, contained HPC with its native signal peptide linked to the P-35S promoter, the extensin signal peptide, and the *nos* poly A region.

The HPC plant expression vectors containing a plant screenable marker were created by digesting the vector at the unique NotI site, which follows the *nos* poly A site, followed by ligation of the GUS (β-glucuronidase) screenable enzyme expression cassette into the

digested vector. The recombinant vectors were analyzed in order to identify vectors having the two plant expression cassettes in the same orientation. The resultant vector pwrg4637 contained HPC with the plant signal peptide (from pwrg4589) and vector pwrg4638 contained HPC with the native HPC signal peptide (from pwrg4590).

Combinations of the genes encoding γ-glutamyl carboxylase, PACE , human HPC, and plant screenable markers, are introduced into plant cells for production of HPC with the desired post-translational modifications.

Plant material propagation: Transformation of competent soy embryos achieved by methods known in the art. McCabe et al., 6 BIO/TECH. 923-26 (1988). GUS screening identified positive transformants.

Transfer of plasmids into *Agrobacterium*: Three plasmids (pwrg4589 encoding HPC, pwrg4069 encoding human γ-glutamyl carboxylase, and pwrg4645 encoding PACE) are purified from *E. coli* using an alkaline lysis DNA minipreparation. Vector DNA is transferred to *A. tumefaciens* strain LBA4404 using a freeze and thaw method.

Preparation of *Agrobacterium* inoculum: *A. tumefaciens* is grown in Luria Broth (LB) (1% tryprone, 0.5% yeast extract, 85 mM NaCl, pH 7.0) medium supplemented with 50 μg/ml kanamycin and 25 μg/ml of streptomycin for 18 hours or until the optical density at 595 nm reaches 0.5 to 1.0. Cells are spun down at 3,000g for 5 minutes and the pellet is resuspended to its initial volume with MS-104 medium (MS basal salts, B5 vitamin, 3% sucrose, 1.0 μg/ml benzylaminopurine (BAP), 0.1 μg/ml naphtaleneacetic acid (NAA), pH 5.7-5.8, and 0.8% agar).

Leaf disc transformation: Leaf squares of about 64 mm^2 are dissected using a sharp scalpel, immersed in the inoculum for 15-30 minutes and plated onto MS-104 medium for 2 days under a 16-hour photoperiod, under low light intensity (20 μE) at 24°C for co-cultivation. Leaf discs are washed alternately three times with sterile distilled water for 1 minute and sterile distilled water supplemented with 500 μg/ml of carbenicillin for 5 minutes. Washed leaf discs are transferred to MS-104 medium with 500 μg/ml of carbenicillin for another 2 days under the same environmental conditions.

Selection and regeneration: Leaf discs are washed as above, plated on MS-104 medium with 500 μg/ml of carbenicillin and 100 μg/ml of kanamycin, and grown using the above environmental conditions until calli appear. The light intensity is increased to 50 μE

and the explants are allowed to grow until they develop well-formed shoots. Shoots are excised and transferred onto MS-rooting medium (MS-104 but with 0.6% agar and no plant growth regulators) with 500 μg/ml of carbenicillin and 100 μg/ml of kanamycin. Surviving plantlets with well-formed roots are removed from the artificial medium, dipped alternately in a 0.06% 50WP solution and in a rooting powder (Stim-root #1) containing indole-3 butyric acid, and are transplanted into pasteurized PROMIX® soil mixture. Plantlets are covered with a transparent cover which was gradually lifted during the following 7 days.

Transformation: Leaf discs are inoculated with *Agrobacterium* inocula.

Screening plants for HPC, γ-glutamyl carboxylase, and PACE expression: A DAS-ELISA procedure is preferred to other ELISA methods because this type of assay is less susceptible to non-specific binding of antibodies. Polyclonal antibodies for HPC, γ-glutamyl carboxylase, and PACE are selected for the sandwich complex. The ELISA for HPC, γ-glutamyl carboxylase, and PACE are used to screen all recovered plants engineered with the pwrg4589, pwrg4069, and pwrg4645 vectors, and to determine which plants are potentially highly expressing HPC, γ-glutamyl carboxylase, and PACE. The antibodies used to detect HPC, γ-glutamyl carboxylase, and PACE expression include rabbit anti-HPC, anti-γ-glutamyl carboxylase, and anti-PACE antibodies and goat anti-HPC, anti-γ-glutamyl carboxylase, and anti-PACE antibodies. The dilution of swine anti-goat IgG is 1:3,000 as recommended by the manufacturer. Finally, the concentration of soluble protein in sap extracts is adjusted to 1 mg/ml.

Quantification of HPC expression: The results from the first ELISA screening are confirmed, and the amount of HPC in transformed plants is estimated by a second ELISA assay using the above antibodies and sample dilutions. Sap is extracted from each plant selected during the screening, and triplicates of the samples are incubated with the antibodies. The best plants are selected for a final quantification of HPC and verified for the expression of the GUS marker gene. The percentage of HPC relative to plant proteins is determined, as well as the standard deviation.

Biological activity using delay in coagulation time: The ACTICLOT® assay kit (American Diagnostica, Greenwich, CT) is used to determine if the plant produced recombinant HPC is biologically active.

Prior to testing, HPC samples are prepared as follows. A 50 µl volume of prepared sample (50 µl undiluted sample plus 400 µl American Diagnostica's dilution buffer) is mixed with an equal amount of HPC deficient plasma and incubated for 2 minutes. A volume of 50 µl of Acticlot activator is then mixed with the sample solution and incubated for an additional five minutes. Finally, 50 µl of calcium chloride stock solution is added and clotting time is monitored by the tilt-tube technique.

Example 6: Production of Post-Translationally Modified Collagen in Dicot Cells

Several regions within a gene sequence such as the 5' untranslated region (5'UTR), translation start context, signal peptide, and 3' untranslated region (3'UTR) can modify gene expression. ALBERTS ET AL., MOLECULAR BIOLOGY OF THE CELL (Garland Pub., Inc., New York 1994). To test the effect of variations of these sequences on the expression of collagen, P4Hα, and PH4β, plant expression vectors were constructed to compare the length of the 3'UTR and the 5'UTR to the signal peptide segment. Generally, the length of the 3'UTR is relatively long, thus truncations of this region would allow for smaller gene cassettes and potentially increase transformation success. The truncations would also eliminate any foreign DNA sequences that may reduce transcription or transcript stability.

The effects of plant and mammalian signal peptides on transcription, translation, and secretory efficiency were also analyzed. As with most secretory proteins, the signal peptide is cleaved from the protein as it is imported through the secretory pathway. Therefore, to ensure proper protein processing and function, the plant signal peptide sequences were ligated in-frame to maintain the coding region of the protein. The ligation site was determined either from published data on the sequenced proteins, or deduced by examining the amino acid sequence and hydrophobicity plot of the proteins. The deduced amino acid sequences at the point of ligation for each protein is shown in Figure 22. The various cDNAs were cloned using molecular biology techniques that are well known in the art. (Sambrook et al. and Ausabel et al., *supra*). A summary of the various vectors can be found in Figure 18.

The pwrg4704 vector contains the prolyl 4-hydroxylase β cDNA ("P4Hβ") (Vuori et al., 89 P.N.A.S. 7467-70 (1992)), with a 35 bp 5' UTR and a 204 bp 3' UTR, and the plant cauliflower mosaic virus 35S promoter (P35S) and the *Agrobacterium tumefaciens* nopaline synthase 3' polyadenylation sequence (NOS 3').

The pwrg4708 vector contains the P4Hβ cDNA, but further truncates the 3'UTR to a 19 bp fragment, within the context of P35S and NOS 3.'

The pwrg4721 vector contains the 3' truncated P4Hβ cDNA, P35S, and NOS 3' as in the pwrg4708, but replaces the native signal peptide with the plant *Nicotiana plumbaginofolia* extensin 5' UTR and signal peptide (ExtSP; Francisco et al., 8 BIOCONJUG. CHEM. 708-13 (1997)).

The pwrg4700 vector contains the prolyl 4-hydroxylase α gene ("P4Hα") cDNA (Vuori et al., 89 P.N.A.S. 7467-70 (1992)) with a 67 bp 5' UTR and a 556 bp 3' UTR within the context of the P35S promoter and NOS 3' terminator.

The pwrg4709 vector contains the P4Hα cDNA within the context of P35S and NOS 3', as in pwrg4700, but further truncates the 3' UTR to a 120 bp fragment.

The pwrg4713 vector contains the 3'UTR truncated P4Hα cDNA, as in pwrg4709, but replaces the native signal peptide with the ExtSP.

The pwrg4716 vector contains both the P4Hα cDNA and the P4Hβ cDNA with their native signal peptides and truncated 3'UTRs.

The pwrg4724 vector also contains both the P4Hα cDNA and the P4Hβ cDNA with the ExtSP signal peptide and truncated 3' UTRs.

The pwrg4710 vector contains the human collagen type III cDNA (Lamberg et al., 271 J. BIOL. CHEM. 11988-95 (1996)) with a 15 bp 5' UTR and 55 bp 3'UTR within the context of P35S and NOS 3'.

The pwrg4715 vector combines the native SP collagen cassette with the plant selectable marker (P35S-aph-NOS 3').

The pwrg4712 vector contains the human collagen type III cDNA as in pwrg4710, but replaces the native signal peptide with the ExtSP.

The pwrg4723 vector combines the ExtSP collagen cassette with the plant selectable marker (P35S-aph2-NOS 3').

A biolistic system was used for gene delivery. Christou et al., 1991, 9 BIO/TECHNOLOGY. 957-61 (1991). DNA was precipitated onto gold beads by suspending 10 mg 0.71 μm gold beads and plasmid DNA in 50 μl buffer (10 mM Tris-Cl (pH 8), 150 mM NaCl), 75 μl 0.1 M spermidine, 75 μl 25% PEG 3500, and 75 μl 2.5 M $CaCl_2$. After incubating for 20 minutes, the DNA-gold bead mixture was briefly centrifuged, rinsed twice

with ethanol, and resuspended in 10 ml ethanol. The mixture (160-320 µl) was then layered onto a mylar sheet, allowed to settle 1-5 minutes, drained, and dried. The mylar sheet was then placed in a biolistic device under reduced atmosphere and by electric discharge, the gold beads were embedded within the plant tissue.

The plant tissue was placed on agar-based media plates, to provide an adhesion surface for the tissue during gene delivery, as well as moisture and nutrients to maintain tissue health during gene expression. Depending on the plant species and tissue, various nutrients including 1-5% sucrose or other sugar, plant macro- and micro- nutrients as in Murishage and Skoog salts; plant growth regulators such as auxins and cytokines; and antibiotics to minimize contamination of the plant material may be used in the culture media.

Plant cells were extracted by grinding the tissue to disrupt the cells, and in some cases, sonication was also utilized to increase the recovery of the desired proteins from the cellular components. Different extraction buffers could be successfully used: Buffer 1: 10 mM Tris (pH 8), 100 mM NaCl, 100 mM glycine, 10 µM DTT, 0.1% Triton-X 100, 1 µg/ml leupeptin, 0.4 mM PMSF; Buffer 2: 2 M Tris (pH 7.5), 5 M NaCl, 1M DTT, 10% Triton-X 100, 1 mg/ml leupeptin, 25 mM PMSF; Buffer 3: 2% SDS, 50 mM Tris (pH 6.8), 10% glycerol, 0.25 or 5 mM EDTA, 1 µg/ml leupeptin, 0.2 mM PMSF.; Buffer 4: 50 mM Tris (pH 8), 0.4 M sucrose, 100 mM KCl, 20 mM $MgCl_2$, 1 mM DTT, 2 µM leupeptin, 0.25 mM PMSF. *See* Bolwell et al., 229 Biochem. J. 693-99 (1985).

Protein extracts were resuspended in 2% SDS, 50 mM Tris (pH 6.8), and 10% glycerol; for reduction, 2.5% β-mercaptoethanol was added to the samples and then, the samples were heated for 10-15 minutes at 70-80°C. The proteins were separated by SDS-PAGE on Novex 4-20% polyacrylamide gels and transferred to 0.45 µm nitrocellulose in 12.5 mM Tris/96 mM glycine buffer at room temperature. Nitrocellulose blots were blocked in 2% nonfat milk, 1% BSA, 100 mM Tris (pH 7.5), and 0.05% Tween 20, then incubated with a primary antibody in blocking buffer. The primary antibodies used in these experiments included anti-collagen propeptide (1:1000 to 5000; Chemicon AB764); anti-collagen type III (1:5000: Biodesign International T59105); anti-P4Hα monoclonal antibody (1:2000; FibroGen, lot 3C4/2F7 012996); anti-P4Hβ (1:1000; DAKO M0877); or anti- P4H rabbit sera (1:2000 to 1:10,000; FibroGen). The blots were then washed three times (3x) in Tris/Tween buffer, followed by incubation with the appropriate secondary antibody (i.e., conjugated to either horseradish peroxidase or alkaline phosphatase) in blocking buffer, washed, followed

by chemiluminescent detection for horseradish peroxidase (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) or 5-bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (BCIP/NBT) detection for alkaline phosphatase (Sigma). The protein standard was 0.1 µl of baculovirus expressing P4Hα, or an P4Hα+β coexpressor (from FibroGen).

To determine whether mammalian genes could be successfully expressed in plant cells, the transient expression of the P4H subunits was demonstrated in soy plant tissue. The results of this study showed that P4Hβ subunits could accumulate in soy plant tissue, with or without the co-introduction of P4Hα.

The growing hypocotyl of soy seed was removed and plasmid DNA was transferred to the target tissue via a biolistic device. Protein extracts were isolated two to four days after gene delivery. As described in step 2, protein was extracted using Buffer 3 and the P4H subunits were detected using the anti-P4H rabbit sera and either the horse radish peroxidase or alkaline phosphatase detection systems

Stable transformation of BY2 tobacco cells was also utilized to confirm the expression of mammalian genes in plants, and to develop material for the purification of collagen. In addition, the enzymatic activity of P4H and its effect on the coexpressed collagen were analyzed in these stably transformed plant cells.

Because the collagen gene is much larger than any previously transferred to plant cells, initially the cells were only transformed with the collagen gene and a selectable marker to assess whether protein expression of collagen was feasible. In later experiment, plant cells were co-transformed with both the collagen/selectable marker plasmid and the P4H plasmid.

DNA preparation and delivery were achieved as described and cell handling and selection were performed as described in Francisco et al., 8 BIOCONJUG. CHEM. 708-13 (1997). BY2 tobacco cells were maintained at 28°C in the dark with shaking (150 rpm). The cells were then plated, bombarded, and rested for one day. Following transformation, the cells were transferred to reduced osmoticum media (0.1 M mannitol, 0.1 M sorbitol) for one day, and then resuspended in media containing 350 µg/ml kanamycin. Individual growing foci were selected and grown on agar plates prior to screening.

An analysis of RNA was performed by Northern blot to assess the transfer of the collagen gene cassette to stably transformed cells. RNA was extracted from the transformed plant cells by the hot phenol method (Ausabel et al., *supra*). Approximately one gram of plants cells were powdered under liquid nitrogen, and then mixed with 3.3 ml of buffer (180

mM Tris (pH 8), 4.5 mM EDTA, 90 mM LiCl, 1% w/v SDS, and 120 ml phenol). Following homogenization by polytron, 3.3 ml of chloroform/isoamyl alcohol (24:1) was added to the cells, and then the cells were vortexed and heated for 20 minutes at 50°C. The aqueous supernatant was precipitated with 0.3 M sodium acetate and 0.6 volumes isopropanol at -20° C overnight. The precipitate was solubilized with 1 ml TE, and then repeatedly extracted with phenol/chloroform/isoamyl alcohol (25:24:1) until most of the interface was gone. The aqueous layer was then precipitated with 1/3 volume 8 M LiCl, at 4°C, and centrifuged. The pellet was rinsed in 2 M LiCl and resuspended in 200 µl TE. Separation by formaldehyde/agarose electrophoresis, transfer to nylon membranes, probe preparation, and hybridization/wash conditions were performed as described by Russell & Sachs, 1 PLANT CELL. 793-803 (1989). The gel contained 0.9% agarose, 9% formaldehyde, 40 mM MOPS, 10 mM sodium acetate, 1 mM EDTA, and 10 ng/ml ethidium bromide. The probe was a 958 bp fragment of human collagen type III isolated from pwrg4710 and randomly labeled with P^{32}. Following electrophoresis and transfer, the RNA blot was incubated with the radiolabeled probe in hybridization buffer (0.25 M $NaPO_3$ (pH 7.2), 5% SDS, 0.1 mg/ml denatured salmon sperm DNA, 1 mM EDTA, and 1% w/v BSA) at 65°C. Following hybridization, the blot was washed with a wash buffer (0.04 M $NaPO_3$ (pH 7.2), 5% SDS, 1 mM EDTA, and 0.5% w/v BSA) at 65°C, and then a final wash in buffer containing 0.25 M $NaPO_3$ (pH 7.2), 1% SDS, 1 mM EDTA, followed by detection on X-ray film.

An example of this Northern blot analysis is shown in Figure 19. Effective integration and expression of the 4.6 Kb collagen coding sequence was observed in several transformed plant cells which are denoted by the underlined sample numbers in Figure 19. A smaller band which may indicate a possible truncated insertion was also detected in some samples. A lack of detectable RNA signal probably indicates a noneffective integration of the gene cassette.

To determine if the co-integration of different genes or the use of different signal peptides influenced expression, transformant calli were created by combining different gene cassettes. First, the expression level, mobility, and truncation fragments of the P4Hβ protein, relative to a standard, were analyzed via Western blot using rabbit sera. Subsequently, the expression of collagen and P4Hα were also evaluated via Western blot using a P4Hα monoclonal antibody and a collagen antibody. Of 32 calli transformed with pwrg4716 (P4Hα

and P4Hβ with native SP) and pwrg4715 (collagen with native SP), 16 scored positive for the P4Hβ protein. Sixteen of 40 calli transformed with pwrg4716 (P4Hα and P4Hβ with native SP) and pwrg4723 (collagen with ExtSP) were positive for the P4Hβ protein. Of 30 calli transformed with pwrg4724 (P4Hα and P4Hβ with ExtSP) and pwrg4723 (collagen with ExtSP), 21 screened positive for the P4Hβ protein. These results indicated that the vectors were successfully expressed, but the plant ExtSP may enhance expression. Following the P4Hβ protein screen, plant cell lines were screened for P4Hα, and then collagen. The number of expressers found for each transformation set and the number screened for that trait (in parentheses) are shown in Figure 23. The expression of P4Hα appears to benefit from the use of the ExtSP.

Positive calli were transferred to liquid media under the conditions described in step 4, but kanamycin was included in the media to maintain selection of transgenic cells. Cells were then subcultured into fresh media at weekly intervals to maintain cell growth. To determine time of harvest for optimal product yield, cell samples were collected throughout the growth cycle period. Cell lines were also evaluated to ensure continuous expression of the three proteins. After 10 days of culture, cells were collected and protein was extracted in one volume of Buffer 1 by grinding the cells with a mortar and pestle, and a small amount of glass beads. The proteins were separated by SDS-PAGE under reducing conditions, followed by transfer to nitrocellulose, and immunodetection for collagen (Chemicon antibody), P4Hα (monoclonal), and P4Hβ (DAKO).

As demonstrated in Figure 20, most cell lines maintained expression of all three genes. Lanes 1-5 represent pwrg4723/4724 cell cultures; lanes 6-7 represent pwrg4715/4716 cultures; and lanes 8-9 represent pwrg4723/4716 cultures. Lane 10 is null, and lanes 11-13 represent a dilution series of purified control standards of these proteins.

Collagen may be directed through the secretory pathway by the secretion signal and accumulate in the media. Failure to accumulate may be due to limitations in passing through the cell wall or secretory compartments within the cell. In order to perform its enzymatic function, P4H is assembled mainly within the cell.

To determine if the proteins accumulate cell-free, cells were grown as described above, and an aliquot of suspension was collected and centrifuged to separate the media from the cells. The media and soluble portion of the cell extract were then applied directly to the

gel (SDS-PAGE) under reducing conditions. The proteins were transferred to nitrocellulose followed by detection of collagen (Chemicon antibody) and P4Hβ (sera). Protein expression was still observed in the cell extract for both proteins. In the media, however, collagen was not detected and P4Hβ had a relatively low signal. The results of this experiment are shown in Figure 21, lanes 1-6 represent pwrg4723/4724 transformant; lanes 7-12 represent pwrg4715/4716 transformant cultures; lane 13 represents a pwrg4723/4716 transformant culture; and lane 14 represents the protein standards.

Northern blot analysis demonstrated the successful integration of the collagen cassette and transcription of the 4.6 kb RNA in plant cell culture suspensions. Furthermore, the expression of the mature proteins, e.g., human collagen type III and the post-translational modifying enzymes, P4Hα and P4Hβ, were confirmed by Western blot analysis.

Example 7: Expression of Immunoglobulins with Mammalian-like Glycosylation in Potato Plants

In a further approach, bovine GalT and GalT derivatives with exchanged CMS domains, are placed in an expression vector for production in potato tubers only. The plasmids pMON415POT 1 to 4 are created from the patatin promoter expression cassette pMON41508, by introducing the GalT- or GalT derivative-encoding sequences. The assembled genes are then transferred to an *A. tumefaciens* expression cassette, pMON34018 for plant transformation.

More specifically, the sequence of bovine GalT has been previously described by D'Agostaro, *supra*. The pMON415POT plasmids place the expression of bovine GalT and derivative sequences, generated by RT-PCR of bovine liver polyA$^+$ RNA as described above, under the control of the patatin promoter. Rosahl et al., 203 MOL. GEN. GENET. 214-20 (1986). Additionally, other regulatory elements of PMON415POT 1-4 include tobacco extensin leader sequence and the soybean RUBISCO SSU 3' UTR. The plasmid selection marker is kanamycin.

Potato cultivars are made competent for transformation, and undergo *Agrobacterium*-mediated transformation using glyphosate selection and regeneration of potato plants as described in U.S. Pat. No. 4,970168.

As another specific example, potato shoot cultures are grown *in vitro* on MS basal medium containing 2.5% sucrose, 0.1 mg/liter indoleacetic acid, and 0.6% agar at 23 ± 2°C in

a 16-h photoperiod. The internodal stem segments (3-5 mm) are incubated for 30 minutes in a saturated culture of *Agrobacterium tumefaciens* containing the pMON415POT expression plasmid, blotted dry on sterile Whatman paper, and transferred onto PR (potato regeneration) medium for 24 hours at 23°C under a 16-hour photoperiod. The PR medium is a typical MS basal medium containing 2.5% sucrose, 1.5 mg/liter benzylaminopurine, 0.1 mg/liter naphthaleneacetic acid, and 0.1 mg/liter gibberellic acid (pH 5.8), solidified with 0.8% agar. The co-cultivation continues for 2 days, and the calli are initiated by transferring the explants onto PR medium containing 250 mg/liter cefotaxime and incubated for 1 week. To obtain a rapid growth of calli, two to three transfers are made onto fresh PR medium. Shoot-buds are initiated on selective and regenerative medium, PRS (PR medium containing 250 mg/liter cefotaxime and 100 mg/liter kanamycin). Kanamycin-resistant shoot-buds (5-6 mm) were transferred onto rooting medium (MS basal medium containing 2.5% sucrose, 0.2 mg/liter indoleacetic acid, solidified with 0.6% agar) supplemented with 100 mg/liter kanamycin.

DNA from transgenic plantlets are isolated and the approximate copy number of the GalT sequence and the site of integration is detected by Southern blot analysis. The intactness of the GalT gene in transgenic plants is determined by PCR using the appropriate GalT-specific primers. Total RNA is extracted from leaves, stems, and tubers of pMON415POT plants using the SNAP kit (Invitrogen Corporation, Carlsbad, CA) following the manufacturer's instructions. The reverse transcription (RT) is done by using GeneAmp EZ rTth RNA PCR kit (Perkin-Elmer). The RT-PCR is carried out with 20 pmol each of GalT gene-specific primers, 300 μM each dNTP, 2.5 mM $Mn(OAc)_2$, 2.5 units of rTth DNA polymerase, and 1 μg of template RNA in a 25-μl reaction volume. The cDNA is then synthesized by PCR.

Transgenic plants are analyzed for the expression of GalT by performing Western blotting with a GalT specific antibody. Total soluble protein is extracted from 250 mg of tissue of transgenic tubers in 0.5 ml of extraction buffer (25 mM Tris·acetate (pH 8.5), 0.5 M NaCl, 5 mM PMSF). The homogenate is centrifuged at 12,000 × *g* for 10 minutes. Protein concentration in the supernatant is measured by using the Bio-Rad assay protein kit (Bio-Rad Laboratories, Hercules, CA). Proteins are separated on SDS-PAGE (15%) gels by using the Mini-Protean II system (Bio-Rad Laboratories, Hercules, CA). The gels are stained with Coomassie brilliant blue R250 to determine protein quality and quantity.

For Western blotting, the SDS/PAGE-separated proteins are transferred onto a nitrocellulose membrane (Schleicher & Schuell, Inc., Keene, NH). The blots are probed with antibodies as described in Example 4. Polyclonal antibodies are diluted using 0.5% BSA in PBS, and binding of polyclonal antibodies is detected with anti-rabbit Ig-conjugated alkaline phosphatase (Promega, Madison, WI). An immunoreactive protein is detected from the transgenic plants.

Moreover, GalT activity is assayed by using the method as set forth in Example 1. The amount of radiolabeled UDP is measured using a liquid scintillation counter. The enzymatic activity is presented as units of heterologous polypeptide per μg of total soluble protein.

After GalT activity has been successfully detected, the GalT transgenic potato plants are cross-pollinated with potato plants expressing a heterologous polypeptide for which altered glycosylation is desired. For example, human monoclonal antibody, such as huNR-LU-10 has been successfully expressed in potato tubers. The huNU-LR-10 transgenic potato plants were transformed with the huNU-LR-10 plasmid as described above. Total soluble protein extracts from the NU-LR-10 transgenic plants were prepared and Western blots performed, as described above, to detect the presence of recombinant antibodies. The detection was carried out with anti-heavy chain antibodies and peroxidase conjugation (Figure 24). Potatoes expressing single-chain antibody may be used in an alternative approach.

After the GalT- and heterologous polypeptide-expressing plants are cross-pollinated, potatoes are grown to maturity and then harvested. Several potatoes are analyzed, by standard techniques, as described in previous examples, for altered glycosylation of the target molecules produced therein.

Example 8: Expression of Immunogobulins with Mammalian-like Glycosylation in Soybean Plants

In addition to corn, tobacco, and potato, the production of antibodies with mammalian-like glycosylation can also be accomplished in soybean. Antibodies have been produced in soybean plants, as described by, e.g., Tremaine et al., ANNUAL MEETING OF THE AMERICAN SOCIETY OF PLANT PHYSIOLOGISTS (June 1988), Schlittler et al., 13H SYMP. PROTEIN SOC'Y, POSTER 588-M (July, 2000), and Bassuner et al., S-28-3 6TH INT. CONGRESS

PLANT MOL. BIOL. Quebec, Canada (2000). The glycan structure of soy plant-produced antibodies have also been described. (McCormick et al., 5 PLANT CELL REPORTER 81-84 (1986).

The plasmids pMON415SOY 1 through 4 are created from pMON41522, by exchanging the glutelin promoter and the HSP70 intron for a constitutive promoter, such as the 35S CaMV (McCormick et al., *supra*). Additionally, the pMON415SOY 1-4 constructs include the tobacco extensin 5' leader sequence and the soybean RUBISCO SSU 3'UTR. The assembled genes are then transferred into an expression cassette, pMON33510, for plant transformation.

The GalT expression cassette is transformed into soybean embryos by particle bombardment as described in U.S. Pat. No. 5,914,451. As a specific example, to induce somatic embryos, cotyledons (3-5 mm in length) are dissected from surface sterilized immature soybean seeds of soybean cultivar. Secondary somatic embryos to be used for bombardment are excised and maintained essentially as described in FINER & MCMULLEN, 27 IN VITRO CELL. DEVEL. BIOL. (1991). The embryonic culture is preferably placed under low intensity lighting. Once the embryos have formed, e.g., after about 15 to 30 days, they are moved to a separate maturation medium, but preferably are kept on the embryogenesis medium until mature, *i.e.* at about 30 days, or when at least about 2.5 mm long. Selection of transformants is achieved by the addition of kanamycin to the media. The transformed explants are then allowed to regenerate. Preferably, they are subcultured at biweekly intervals on a regeneration medium containing the antibiotic and selection agent, until shoot apices are formed, then transferred to appropriate shooting, rooting, and soil media. Tissue culture of various tissues of soybeans and regeneration of plants therefrom is well known in the art and widely published. Komatsuda, et al., 31 CROP SCI. 333-337 (1991); Stephens et al., 82 THEOR. APPL. GENET. 633-635 (1991).

Transgenic plants are analyzed for the expression of GalT by performing Western blotting with a GalT specific antibody. Total soluble protein is extracted from 250 mg of tissue of transgenic tubers in 0.5 ml of extraction buffer (25 mM Tris·acetate (pH 8.5), 0.5 M NaCl, 5 mM PMSF). The homogenate is centrifuged at 12,000 × *g* for 10 minutes. Protein concentration in the supernatant is measured by using the Bio-Rad assay protein kit (Bio-Rad Laboratories, Hercules, CA). Proteins are separated on SDS/PAGE (15%) gels by using the

Mini-Protean II system (Bio-Rad Laboratories, Hercules, CA). The gels are stained with Coomassie brilliant blue R250 to determine protein quality and quantity.

For Western blotting, the SDS/PAGE-separated proteins are transferred onto a nitrocellulose membrane (Schleicher & Schuell, Inc., Keene, NH). The blots are probed with antibodies as described in Example 4. Polyclonal antibodies are diluted using 0.5% BSA in PBS, and binding of polyclonal antibodies is detected with anti-rabbit Ig-conjugated alkaline phosphatase (Promega, Madison, WI). An immunoreactive protein is detected from the transgenic plants.

Moreover, GalT activity is assayed by using the method as set forth in Example 1. The amount of radiolabeled UDP is measured using a liquid scintillation counter. The enzymatic activity is presented as units of heterologous polypeptide per μg of total soluble protein.

After GalT activity has been successfully detected, the GalT transgenic soybean plants are cross-pollinated with soybean plants expressing a target heterologous protein, such as human monoclonal antibody as described above. Alternatively, soybean expressing single-chain antibody may be used. After cross-pollination, soybeans are grown to maturity and then harvested. Several soybeans are analyzed, by standard techniques, as described in Example 2, for altered glycosylation of the target heterologous polypeptides produced therein.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

The disclosures of all references and publications cited above are expressly incorporated by reference in their entireties to the same extent as if each were incorporated by reference individually.

We claim:

1. A method for producing a post-translationally modified heterologous polypeptide in a plant host system comprising the step of altering the natural post-translational modification abilities of said plant host system.

2. The method of claim 1 further comprising the steps of:

transforming said plant host system with one or more nucleic acid sequences encoding said heterologous polypeptide; and

isolating said heterologous polypeptide from said plant host system.

3. The method of claim 1, wherein said post-translation modification is selected from the group consisting of proteolytic cleavage, glycosylation, phosphorylation, methylation, sulfation, prenylation, acetylation, N-amidation, oxydation, hydroxylation and myristylation.

4. The method of claim 1, wherein said altering step comprises transforming said plant host system with one or more nucleic acid sequences encoding a post-translational modification enzyme.

5. The method of claim 4, wherein said post-translational modification enzyme includes enzymatically active fragments thereof and is one or more enzymes selected from the group consisting of glycoprotein glycosyltransferases, GlcNAc-1-phosphotransferase, GlcNAc 1-phosphodiester-N-acetylglucosaminindase, glycosidases, exoglycosidases, endoglycosidases, GlcNAc phosphotransferase, protein kinases, 3'-phosphoadenosyl-5'-phosphosulphate, prolyl hydroxylase, and lysyl hydroxylase.

6. The method of claim 4, wherein said post-translational modification is glycosylation.

7. The method of claim 1, wherein said heterologous polypeptide is selected from the group consisting of immunoglobulins, integrins, addressins, selectins, homing receptors, T-cell receptor units, soluble major histocompatibility complex antigens, growth factor receptors,

growth factors, growth hormones, cell cycle proteins, viral antigens, bacterial antigens, vaccines, fibrinogen, thrombin and hyaluronic acid.

8. The method of claim 1, wherein said heterologous polypeptide comprises a blood protein selected from the group consisting of serum albumin, hemoglobin, Factor VII, Factor VIII, modified Factor VIII, Factor IX, Factor X, tissue plasminogen factor, Protein C, von Willebrand factor, antithrombin III, and erythropoietin.

9. The method of claim 1, wherein said heterologous polypeptide is a colony stimulating factor selected from the group consisting of granulocyte colony-stimulating factor, macrophage colony-stimulating factor and granulocyte macrophage colony-stimulating factor.

10. The method of claim 1, wherein said heterologous polypeptide is a cytokine selected from the group consisting of interleukins 1 through 18, interleukin-T, interferon alpha, interferon beta, interferon gamma, leukemia inhibitory factor, oncostatin, transforming growth factor beta, tumor necrosis factor alpha, and tumor necrosis factor beta.

11. The method of claim 1, wherein said heterologous polypeptide is a membrane surface protein selected from the group consisting of insulin receptors, epidermal growth factor receptor, and β-adrenergic receptor.

12. The method of claim 1, wherein said heterologous polypeptide is a structural protein selected from the group consisting of collagen types I through XX, fibrinogen, elastin, tubulin, actin and myosin.

13. The method of claim 1, wherein said heterologous polypeptide is an antibody or a functional equivalent thereof.

14. The method of claim 13, wherein said antibody or functional equivalent thereof, is selected from the group consisting of IgA, IgG, IgD, IgE, IgM, Fab, and Fv.

15. The method of claim 1, wherein said plant host system is a plant.

16. The method of claim 15, wherein said plant is a monocotyledonous plant.

17. The method of claim 16, wherein said monocotyledonous plant is selected from the group consisting of lilies, grasses, corn, orchids, irises, onions, and palms.

18. The method of claim 16, wherein said monocotyledonous plant is a grain selected from the group consisting of oats, wheat, and barley.

19. The method of claim 15, wherein said plant is a dicotyledonous plant.

20. The method of claim 19, wherein said dicotyledonous plant is selected from the group consisting of tobacco, tomato, alfalfa, oaks, maples, roses, mints, squash, daisies, potato, soybean, walnuts, cacti, violets, petunias, and buttercups.

21. The method of claim 1, wherein said plant host system is a plant cell suspension culture.

22. The method of claim 21, wherein said plant cell suspension culture is derived from a monocotyledonous plant.

23. The method of claim 21, wherein said plant cell suspension culture is a corn cell suspension culture.

24. The method of claim 21, wherein said plant suspension culture is derived from a dicotyledonous plant.

25. The method of claim 21, wherein said plant cell suspension culture is a tobacco cell suspension culture.

26. The method of claim 1, wherein said altering step comprises performing mutagenesis of plant host system.

27. The method of claim 26, wherein said mutagenesis is targeted to portions of the genome of said plant host system.

28. The method of claim 27, wherein said genomic portion encodes endogenous plant post-translational modification enzymes.

29. The method of claim 1, wherein said altering step comprises transforming said plant host system with an expression vector comprising a nucleic acid sequence that encodes an antisense nucleic acid.

30. The method of claim 29, wherein said antisense nucleic acid sequence inhibits the expression of at least one endogenous plant protein.

31. The method of claim 30, wherein said protein comprise a plant specific post-translational modification enzyme.

32. The method of claim 31, wherein said plant post-translational modification enzyme is selected from the group consisting of GnT1, XylT, and FucT.

33. The method of claim 29, wherein said antisense nucleic acid comprises antisense RNA.

34. The method of claim 29, wherein said antisense nucleic acid comprises antisense DNA.

35. The method of claim 1, wherein said altering step comprising transforming said plant host system with a chimeric nucleic acid sequence.

36. The method of claim 35, wherein said chimeric nucleic acid sequence comprises the N-terminal region of N-acetlyglucosamine transferase gene operably linked to a nucleic acid sequence encoding the GnT1 enzymatic active site.

37. The method of claim 35, wherein said chimeric nucleic acid sequence comprises the N-terminal region of mannosidase I gene operably linked to a nucleic acid sequence encoding the GnT1 enzymatic active site.

38. The method of claim 2, wherein said one or more nucleic acid sequences are contained within one or more expression vectors.

39. The method of claim 38, wherein said expression vector comprises a signal peptide functional in said plant host system, a promoter functional in said plant host system, a selection gene functional in said plant host system, a transcriptional and translational initiation region, and a transcriptional and translational termination region.

40. A method for producing a post-translationally modified heterologous polypeptide in a plant host system comprising the step of expressing said heterologous polypeptide, wherein the cells of said plant host system have been transformed with one or more expression vectors comprising a nucleic acid sequence encoding a heterologous polypeptide.

41. The method of claim 40, further comprising the step of:

transforming said plant host system with one or more nucleic acid sequences, wherein said nucleic acid sequence encodes a post-translational modifying enzyme.

42. The method of claim 40, further comprising the step of isolating said heterologous polypeptide from said plant host system.

43. The method of claim 41, further comprising the step of isolating said hetrologous polypeptide from said plant host system.

44. A method for producing a post-translationally modified heterologous polypeptide comprising the step of expressing said post-translational modifying enzyme, wherein the cells of the plant host system have been transformed with an expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

45. The method of claim 44, further comprising the steps of:

transforming said plant host system with one or more expression vectors comprising a nucleic acid sequence encoding said heterologous polypeptide; and

isolating said heterologous polypeptide.

46. A method for producing a post-translationally modified heterologous polypeptide, wherein the cells of the plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme, comprising the steps of:

expressing said heterologous polypeptide; and

expressing said post-translational modifying enzyme.

47. The method of claim 46, further comprising the step of isolating said heterologous polypeptide from said plant host system.

48. A method for producing a post-translationally modified heterologous polypeptide, wherein the cells of the plant host system have been transformed with one or more expression vectors comprising a first nucleic acid sequence encoding a heterologous polypeptide and a second nucleic acid sequence encoding a post-translational modifying enzyme, comprising the steps of:

expressing said heterologous polypeptide; and

expressing said post-translational modifying enzyme.

49. The method of claim 48, further comprising the step of isolating said heterologous polypeptide from said plant host system.

50. A plant host system expressing a post-translationally modified heterologous polypeptide, wherein the natural post-translational modification abilities of said plant host system have been altered.

51. A plant host system expressing a post-translationally modified heterologous polypeptide, wherein the cells of said plant system have been transformed with an expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide.

52. The plant host system of claim 54, wherein said plant host system is transformed with one or more expression vectors comprising a nucleic acid sequence that alters the natural post-translational modification abilities of said plant host system.

53. The plant host system of claim 51, wherein said plant host system is transformed with one or more expression vectors comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

54. A plant host system expressing a post-translationally modified heterologous polypeptide, wherein the cells of said plant host system have been transformed with an expression vector comprising a post-translational modifying enzyme.

55. The plant host system of claim 54, wherein said plant host system is transformed with one or more expression vectors comprising a nucleic acid sequence wherein said nucleic acid sequence alters the natural post-translational modification abilities of said plant host system.

56. The plant host system of claim 54, wherein said plant host system is transformed with one or more expression vectors comprising a nucleic acid sequence encoding said heterologous polypeptide.

57. A plant host system expressing a post-translationally modified heterologous polypeptide, wherein the cells of the plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

58. A plant host system expressing a post-translationally modified heterologous polypeptide, wherein the cells of the plant host system have been transformed with an expression vector

comprising a first nucleic acid sequence encoding a heterologous polypeptide and a second nucleic acid sequence encoding a post-translational modifying enzyme.

59. A method for producing a post-translationally modified heterologous polypeptide in a plant host system, wherein the cells of said plant system have been transformed with an expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide comprising the step of expressing said nucleic acid sequence.

60. The method of claim 59, further comprising the step of transforming said plant host system with one or more expression vectors comprising a nucleic acid sequence encoding a post-translational modification enzyme.

61. The method of claim 59, wherein said plant host system has been genetically manipulated to alter the natural post-translational abilities of said plant host system.

62. A method for producing a properly post-translationally modified heterologous polypeptide in a plant host system, wherein the cells of said plant host system have been transformed with an expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme comprising the step of expressing said nucleic acid.

63. The method of claim 62, further comprising the step of transforming said plant host system with one or more expression vectors encoding a heterologous protein.

64. A method for producing a post-translationally modified heterologous polypeptide in a plant host system, wherein the cells of the plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme, comprising the steps of:

expressing said nucleic acid encoding said heterologous protein; and

expressing said nucleic acid encoding said post-translational modifying enzyme.

65. A method for producing a post-translationally modified heterologous polypeptide in a plant host system, comprising the step of:

cross-pollinating a first plant host system, wherein the cells of said first plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide, and a second plant host system, wherein the cells of said second plant host system have been transformed with a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

66. The method of claim 65, wherein the resultant plant host progeny from said cross-pollination comprises said first expression vector and said second expression vector.

67. The method of claims 65, further comprising the step of isolating said heterologous polypeptide from said plant host system.

68. The method of claim 66, further comprising the step of isolating said heterologous polypeptide from said plant host system.

69. A method for isolating a post-translationally modified heterologous polypeptide in a plant host system, comprising the step of:

cross-pollinating a first plant host system wherein the cells of said first plant host system have been transformed with a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second plant host system wherein the cells of said second plant host system have been transformed with a second expression vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

70. The method of claim 69, wherein the resultant plant host progeny from said cross-pollination comprises said first expression vector and said second expression vector.

71. A plant host system that produces a post-translationally modified heterologous polypeptide wherein said plant host system expresses a first expression vector comprising a nucleic acid sequence encoding a heterologous polypeptide and a second express vector comprising a nucleic acid sequence encoding a post-translational modifying enzyme.

72. The plant host system of claim 71, wherein said plant host system is the resultant progeny of a cross-pollination between a first plant host system comprising said first expression vector and a second plant host system comprising said second expression vector.

73. A plant produced according the method of claim 1.

74. The seed produced form the plant of claim 73.

75. An expression vector comprising one or more nucleic acid sequences encoding one or more of the following selected from the group consisting of a heterologous polypeptide and a post-translational modifying enzyme.

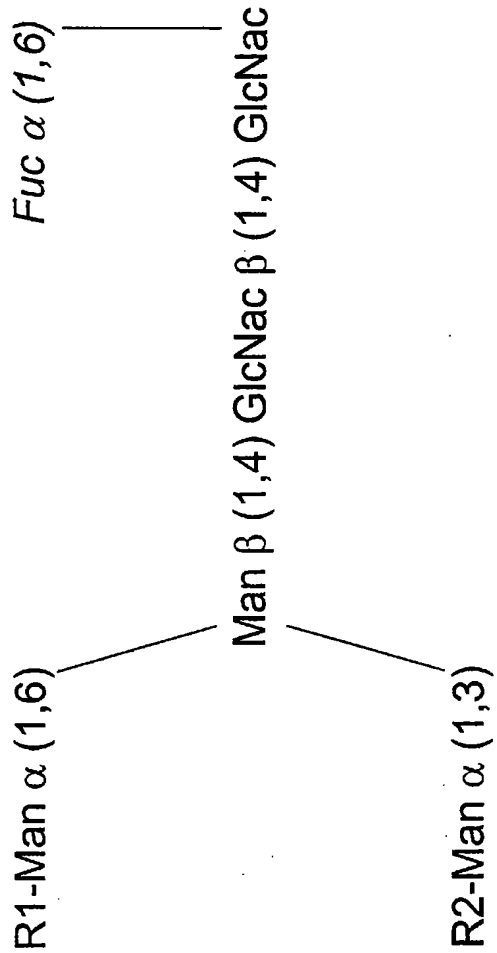


Figure 1

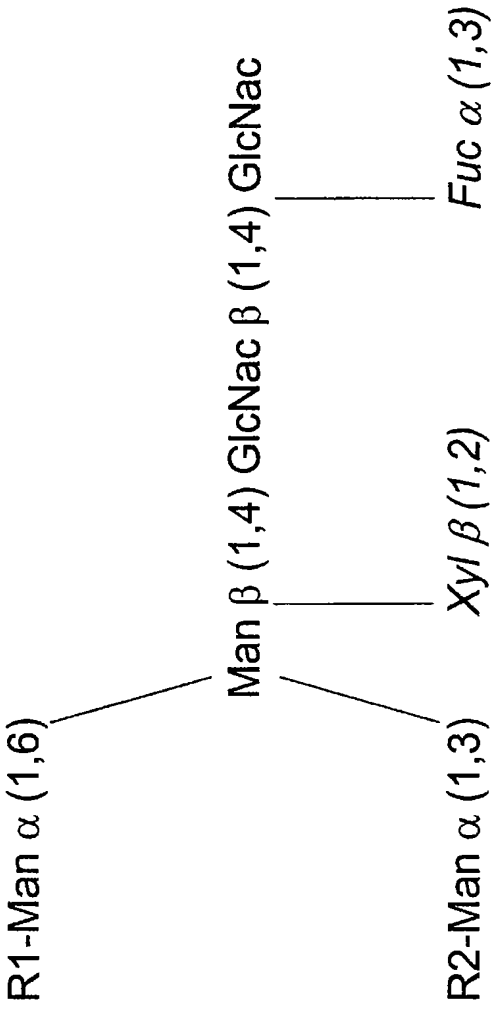


Figure 2

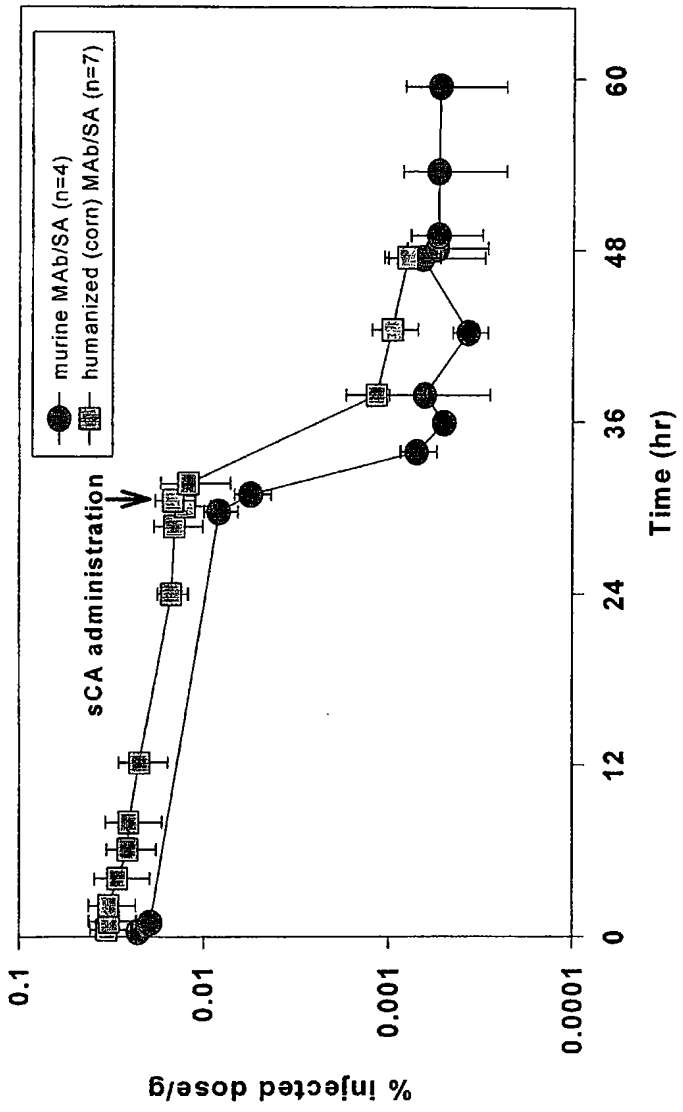

murine MAb/SA (n=4)
humanized (corn) MAb/SA (n=7)
sCA administration
% injected dose/g
0.1
0.01
0.001
0.0001
0
12
24
36
48
60
Time (hr)

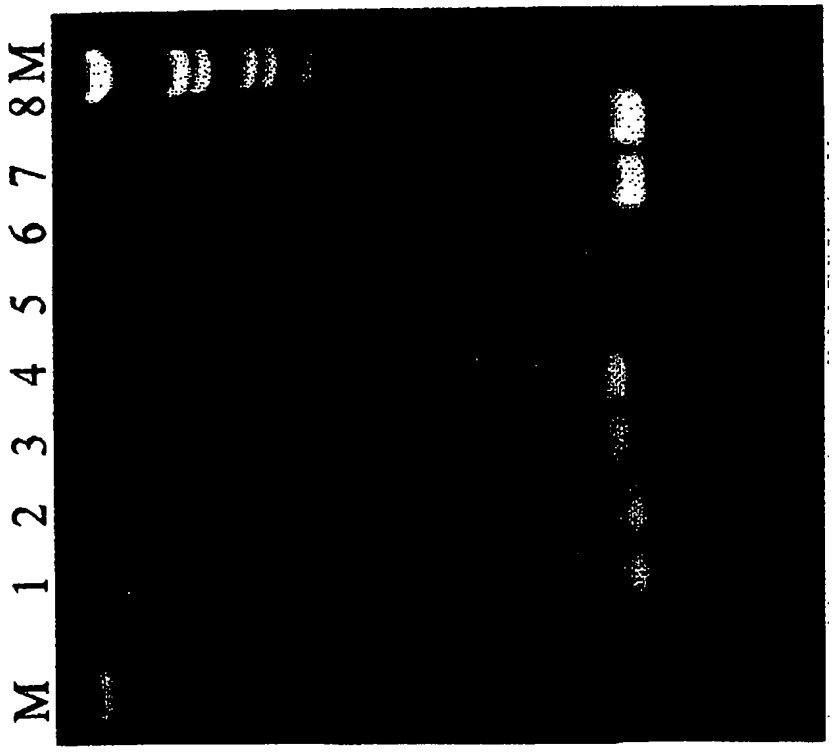
Figure 3

Design of primers for the amplification of mouse, bovine and human GalT through RT-PCR

Primers for GalT of mouse, bovine, and human with XbaI at the 5' and EcoRI at the 3' ends				
The XbaI sites at the 5' would be: CTAGTCTAGACTAG				
The EcoRI sites at the 3' end would be:CCGGAATTCCGG				
Name of the Gene	Primer Number	Primer Sequence	Position of the primers	Product length
Mouse GalT 5'	DAR292	CTAGTCTAGACTAGAGCCCACCCCCTCTTAAAGC	140-159	1229
Mouse GalT-3'	DAR293	CCGGAATTCCGGATGCTATCTCGGTGTCCCGATGTCC	1368-1350	
BovineGalT-5'	DAR294	CTAGTCTAGACTAGGGCGGGAAGATGAAGTTTCG	175-194	1287
BovineGalT-3'	DAR289	CCGGAATTCCGGCAGCACACAGGCAGAGACACAG	1461-1440	
* BovineGalT-5'	DAR295	CTAGTCTAGACTAGGGCGGGACCATGAAGTTTCG	175-194	1287
Human GalT-5'	DAR296	CTAGTCTAGACTAGTAAAGCGGCGGCGGGAAG	58-75	1217
Human GalT-3'	DAR297	CCGGAATTCCGGAAACGCTAGCTCGGTGTCCC	1274-1255	
* Bovine Galt-5'-DAR295 is same as DAR294 except that AAG near the ATG has been replaced with ACC.				
** Position of the primers in corresponding cDNAs of mouse, bovine, and human GalT mentioned in the disclosure.				

Figure 4

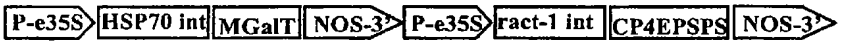
FIG. 5



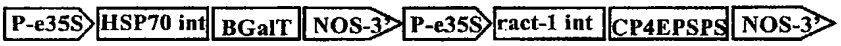
M- Marker
1-2, Mouse GalT
3-4, Bovine GalT
5-6, Bovine GalT with altered primer
7-8, Human GalT

***Zea mays* cell culture vectors**

pMON41411 (contains Mouse GalT cDNA)

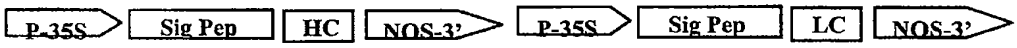


pMON41412 (contains Bovine GalT cDNA)



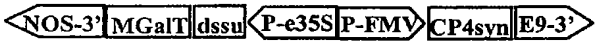
pMON41413 (contains Human GalT cDNA)

pMON41409 (contains the heavy and light chain of huNR-LU-10 Mab)

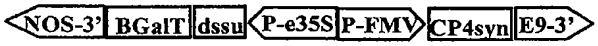


***Nicotiana tabacum* cell culture vectors**

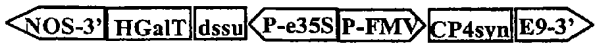
pMON41417 (contains Mouse GalT cDNA)



pMON41418 (contains Bovine GalT cDNA)



pMON41419 (contains Human GalT cDNA)



The antibody construct expressed in tobacco cell lines

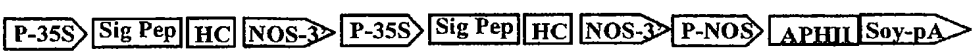
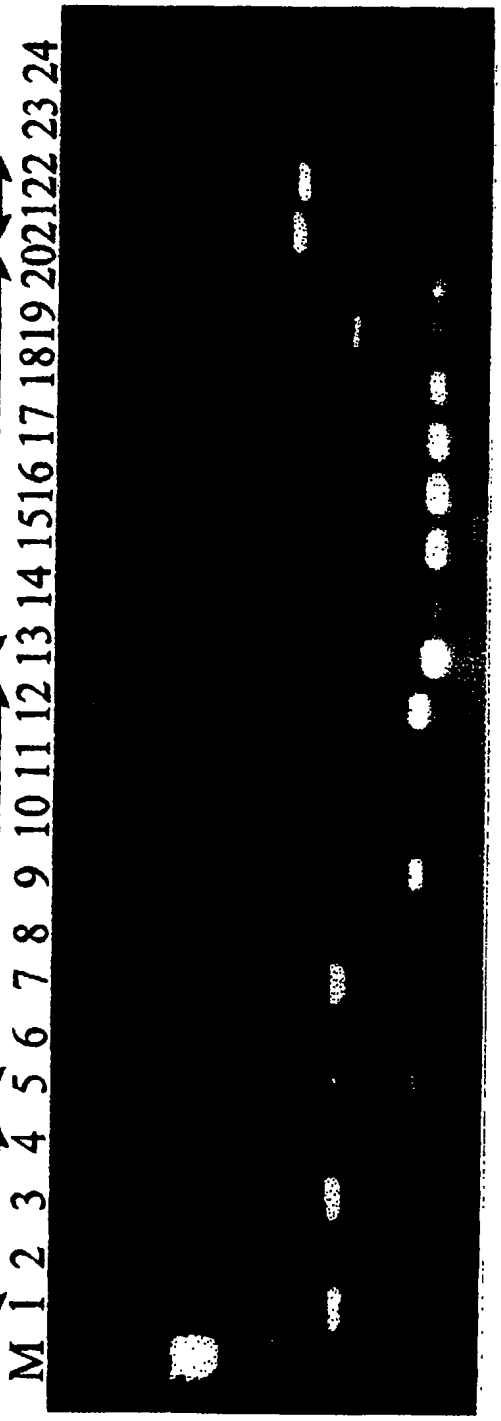


Figure 6

FIG. 7


M 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
MGalT
BGalT
HGalT
HC

Overall positives identified from corn-GalT Northern blot screening:					
Sample ID	Signal	Sample ID	Signal	Sample ID	Signal
2A1-2	strong	3B2-5	strong	4A2-3	strong
2B2-4	strong	3B2-6 (very little)	strong	4A2-4	strong
**2D1-2	strong	3B2-7	strong	*4B1-4	strong
*2B2-6	strong	*3B1-6	strong	4C2-2	weak
2B1-3	Very low	3C1-3	strong	4D2-5	strong
2D1-1	strong	*3B1-4	strong	*4D1-2	strong
2A2-1	strong	*3B2-4	strong	4B1-5	strong
2B1-4	strong	3C1-6	OK	4C2-3	strong
2B2-1	(very weak)			*4A2-1	strong
2B1-4	weak			*4D2-1	strong
2B1-6	weak				

Figure 8

Mouse GalT	Bovine GalT	Human GalT
41417-06	41418-19	41419-03
41417-10	41418-20	41419-04
41417-11	41418-24	41419-13
41417-14	41418-25	41419-17
41417-17	41418-26	41419-24
41417-20	41418-27	41419-25
41417-31	41418-28	41419-35
41417-34	41418-30	41419-36
41417-35	41418-31	41419-39
41417-36		

N.t.-GalT Positives from Northern Blot Analysis

Figure 9

Mouse GalT	ng NRX/ ug TSP	Bovine GalT	ng NRX/ ug TSP	Human GalT	ng NRX ug TSP
2A1-2	1.40	3B1-2	2.79	4A1-2	0.01
2B2-4	0.70	3B2-3	0.28	4A1-3	0.00
2A1-3	0.01	3B2-4	0.16	4B1-2	0.02
2B2-3	0.01	3B2-5	1.07	4B1-3	0.01
2D2-2	0.01	3C1-2	0.17	4D2-4	0.02
2D2-3	0.01	3C1-3	0.01	4D2-2	1.27
2A2-1	0.23	3B1-3	0.50	4A1-4	0.04
2B2-5	0.36	3B1-4	0.03	4A2-3	1.63
2B2-6	0.06	3B2-6	1.01	4A2-4	3.41
2D1-2	0.40	3B2-7	0.34	4B1-4	1.35
2D1-5	0.09	3C1-5	0.07	4B2-1	0.99
		3C2-4	0.05	4C2-2	1.88
		3C2-5	0.15	4C2-3	0.20
				4D1-2	1.38

Expression of huNR-LU-10 antibody in transgenic corn-GalT calli

Figure 10

GalT Calculated Based on STD Curve in Corn				
Sample ID	GalT ug/ml	TSP (ug/ul)	Activity	
			(units/mg protein)	
2D1-2	4.6	6.41	7.033	
4B1-4	1	4.49	2.183	
2D1-2 Golgi	37	1.9	190.842	
2A1-2	1.44	1.52	9.284	
2B2-4	0.08	2.43	0.323	
2A2-1	0.18	2.09	0.844	
2D1-1	0.23	1.59	1.418	
2B1-4	0.47	1.99	2.315	
3B1-6	0.27	2.01	1.316	
418-30	1	0.973	10.072	

Figure 11

Summary of transgenic corn-GalTscreening to identify a cell line that expresses mouse, bovine and human GalT

Mouse GalT	ng NRX/ ug TSP	Signal intensity of GalT mRNA	GalT assay mu/mg TSP	Bovine GalT	ng NRX/ ug TSP	Signal intensity of GalT mRNA	GalT assay mu/mg TSP	Human GalT	ng NRX/ ug TSP	Signal intensity of GalT mRNA	GalT assay mu/mg TSP
2A1-2	1.4	Strong	9.284	3B1-2	2.79			4A1-2	0.01		
2B2-4	0.7	Strong	0.323	3B2-3	0.28			4A1-3	0		
2A1-3	0.01			3B2-4	0.16	Strong		4B1-2	0.02		
2B2-3	0.01			3B2-5	1.07	Strong		4B1-3	0.01		
2D2-2	0.01			3C1-2	0.17			4D2-4	0.02		
2D2-3	0.01			3C1-3	0.01	Strong		4D2-2	1.27		
2A2-1	0.23	Strong	0.844	3B1-3	0.5			4A1-4	0.04		
2B2-5	0.36			3B1-4	0.03	Strong		4A2-3	1.63	Strong	
2B2-6	0.06	Strong		3B2-6	1.01	Strong		4A2-4	3.41	Strong	
2D1-2	0.4	Strong	7.03	3B2-7	0.34	Strong		4B1-4	1.35	Strong	2.183
2D1-5	0.09			3C1-5	0.07			4B2-1	0.99		
2B1-1	0			3C2-4	0.05			4C2-2	1.88	Weak	
2B1-2	0			3C2-5	0.15			4C2-3	0.2	Strong	
2B1-3	0	Very low		3B1-5	0			4D1-2	1.38	Strong	
2B1-4	0	Weak	0.47	3B1-6	0	Strong	1.316	4A2-1	0	Strong	
2B1-5	0			3C1-4	0			4A2-2	0		
2B1-6	0	Weak		3C1-6	0	OK		4B1-5	0	Strong	
2D1-1	0	Strong	0.23	3C2-2	0			4D1-1	0		
2D1-3	0			3C2-3	0			4D2-5	0	Strong	
2D1-4	0							4D2-6	0		

Figure 12

pMON41522 (contains Bovine GalT cDNA)

Figure 13

```
MAN_ATGT      M------------------SCPIHPRRLFL--CLL---ISLTFF----          21
MNS285bGalT   MKFREPLLGGSAAM-----PGASLQRACRLLVAVCALHLGVTLVYY----       41
MAN_GMGT      MARGSRSVGSSSSKWRYCNPSYYLKRPKRLAL-LFIVFVCVSFVFWDRQT       49
GNT_GALT      MR----------GYKFCCDFRYL-----LILAA-VAFIYIQMRLFATQS         33

MAN_ATGT      ------V------------------VDPSSQHIEVKKKQMR------          38
MNS285        ------LAGRDLRR----LPQLVGVHPPLQGSSHGAAAIGQPSG------       75
MAN_GMGT      LVREHQVEISELQKEVTDLKNLVDDLNNKQGGTSGKTDLGRKATKSSKDV       99
GNT_GALT      EYADRLAAAIEAENHCTSQTRLLIDQISQQQGR--IVALEEQMKRQ----       77

MAN_ATGT      -------ELRLRGVAPPPPLQNSSKPRSRAPSNLDAYSHPGPGPGPGSNL       81
MNS285        -------ELRLRGVAPPPPLQNSSKPRSRAPSNLDAYSHPGPGPGPGSNL      118
MAN_GMGT      LDDPIDIELRLRGVAPPPPLQNSSKPRSRAPSNLDAYSHPGPGPGPGSNL      149
GNT_GALT      -----DQELRLRGVAPPPPLQNSSKPRSRAPSNLDAYSHPGPGPGPGSNL      122
```


Figure 14a

```
MAN_ATGT      TSAPVPSTTTRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLG      131
MNS285        TSAPVPSTTTRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLG      168
MAN_GMGT      TSAPVPSTTTRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLG      199
GNT_GALT      TSAPVPSTTTRSLTACPEESPLLVGPMLIEFNIPVDLKLVEQQNPKVKLG      172
```

Figure 14b

MAN_ATGT	GRYTPMDCISPHKVAIIIPFRNRQEHLKYWLYYLHPILQRQQLDYGIYVI	181
MNS285	GRYTPMDCISPHKVAIIIPFRNRQEHLKYWLYYLHPILQRQQLDYGIYVI	218
MAN_GMGT	GRYTPMDCISPHKVAIIIPFRNRQEHLKYWLYYLHPILQRQQLDYGIYVI	249
GNT_GALT	GRYTPMDCISPHKVAIIIPFRNRQEHLKYWLYYLHPMLQRQQLDYGIYVI	222

Figure 14c

MAN_ATGT	NQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMNDHNTYRCFS	231
MNS285	NQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMNDHNTYRCFS	268
MAN_GMGT	NQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMNDHNTYRCFS	299
GNT_GALT	NQAGESMFNRAKLLNVGFKEALKDYDYNCFVFSDVDLIPMNDHNTYRCFS	272

Figure 14d

MAN_ATGT	QPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLSINGFPNNYWGWGGED	281
MNS285	QPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLSINGFPNNYWGWGGED	318
MAN_GMGT	QPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLSINGFPNNYWGWGGED	349
GNT_GALT	QPRHISVAMDKFGFSLPYVQYFGGVSALSKQQFLSINGFPNNYWGWGGED	322

Figure 14e

MAN_ATGT	DDIYNRLAFRGMSVSRPNAVIGKCRMIRHSRDKKNEPNPQRFDRIAHTKE	331
MNS285	DDIYNRLAFRGMSVSRPNAVIGKCRMIRHSRDKKNEPNPQRFDRIAHTKE	368
MAN_GMGT	DDIYNRLAFRGMSVSRPNAVIGKCRMIRHSRDKKNEPNPQRFDRIAHTKE	399
GNT_GALT	DDIYNRLAFRGMSVSRPNAVIGKCRMIRHSRDKKNEPNPQRFDRIAHTKE	372

Figure 14f

MAN_ATGT	TMLSDGLNSLTYMVLEVQRYPLYTKITVDIGTPS	365
MNS285	TMLSDGLNSLTYMVLEVQRYPLYTKITVDIGTPS	402
MAN_GMGT	TMLSDGLNSLTYMVLEVQRYPLYTKITVDIGTPS	433
GNT_GALT	TMLSDGLNSLTYMVLEVQRYPLYTKITVDIGTPS	406

Figure 14g

pMON41523 (contains Bovine GalT cDNA with the N-terminus of *N. tabaccum* GnT1)

pMON41587 (contains Bovine GalT cDNA with the N-terminus of *A. thaliana* Man1)

pMON41588 (contains Bovine GalT cDNA with the N-terminus of *Glycine max* Man1)

Figure 15

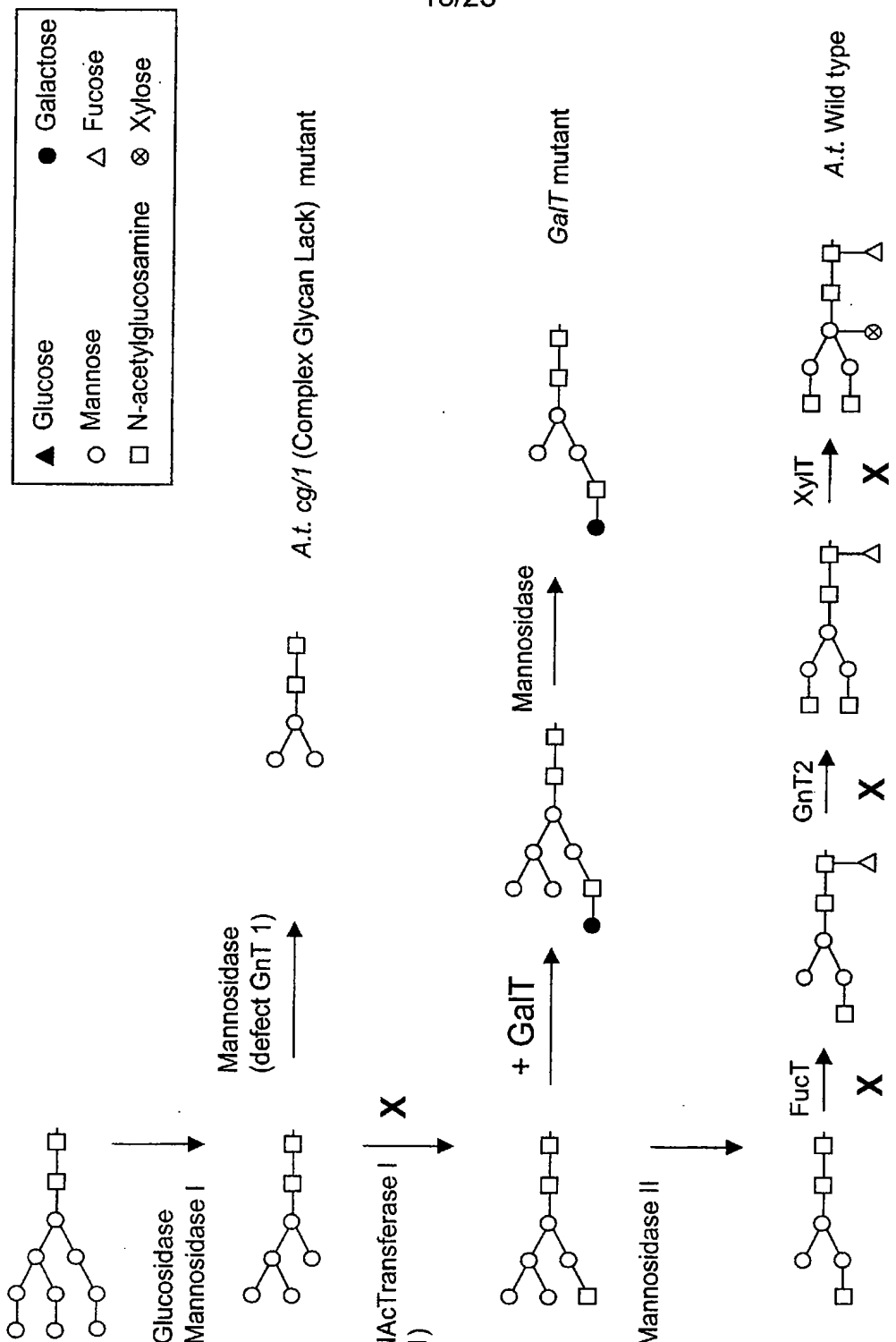

Glucose
Mannose
N-acetylglucosamine
Galactose
Fucose
Xylose
Glucosidase
Mannosidase I
Mannosidase
(defect GnT 1)
A.t. cgl1 (Complex Glycan Lack) mutant
GlcNAcTransferase I
(Gn1)
+ GalT
Mannosidase
GalT mutant
Mannosidase II
FucT
GnT2
XylT
A.t. Wild type

Figure 16

pMON41581 (contains maize GnT1 antisense)

pMON41515 (contains *A. thaliana* XylT antisense)

pMON41582 (contains *Vigna radiaia* FucT antisense)

Figure 17

Plant Expression Vectors			
Vector	UTR		Signal Peptide
Prolyl 4-hydroxylase α			
pwrg4700	67 bp 5' UTR	556 bp 3' UTR	Native
pwrg4709	67 bp 5' UTR	120 bp 3' UTR	Native
pwrg4713	67 bp 5' UTR	120 bp 3' UTR	ExtSP
Prolyl 4-hydroxylase β			
pwrg4704	35 bp 5' UTR	204 bp 3' UTR	Native
pwrg4708	35 bp 5' UTR	19 bp 3' UTR	Native
pwrg4721	35 bp 5' UTR	19 bp 3' UTR	ExtSP
Prolyl 4-hydroxylase α and Prolyl 4-hydroxylase β			
pwrg4716	α: 67 bp 5' UTR β: 35 bp 5' UTR	α: 120 bp 3' UTR β: 19 bp 3' UTR	Native
pwrg4724	α: 67 bp 5' UTR β: 35 bp 5' UTR	α: 120 bp 3' UTR β: 19 bp 3' UTR	ExtSP
Collagen Type III			
pwrg4710	15 bp 5' UTR	55 bp 3' UTR	Native
pwrg4715	15 bp 5' UTR	55 bp 3' UTR	Native and Marker: P35S-aph-NOS 3'
pwrg4712	15 bp 5' UTR	55 bp 3' UTR	ExtSP
pwrg4723	15 bp 5' UTR	55 bp 3' UTR	ExtSP and Marker: P35S-aph-NOS 3'

Figure 18

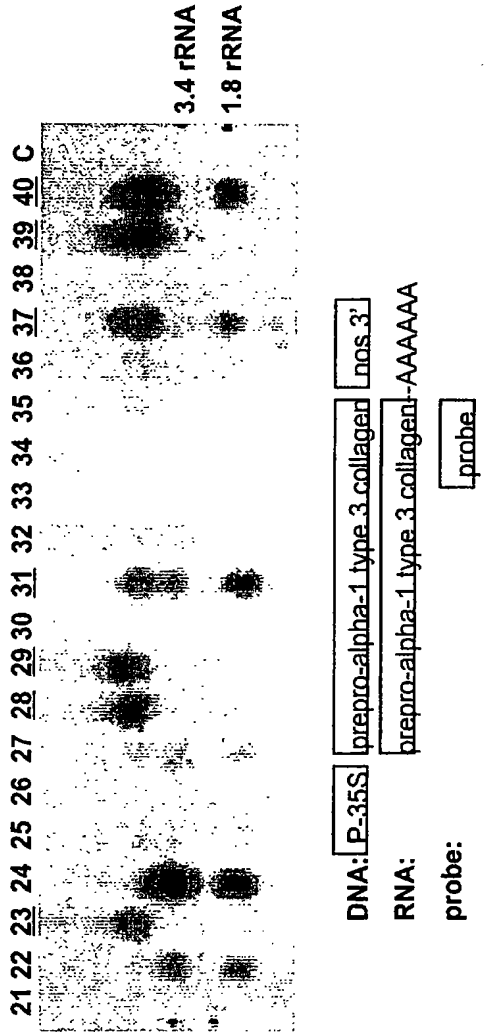

3.4 rRNA
1.8 rRNA
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 C
DNA: P-35S prepro-alpha-1 type 3 collagen nos 3'
RNA: prepro-alpha-1 type 3 collagen -AAAAA
probe: probe

Figure 19

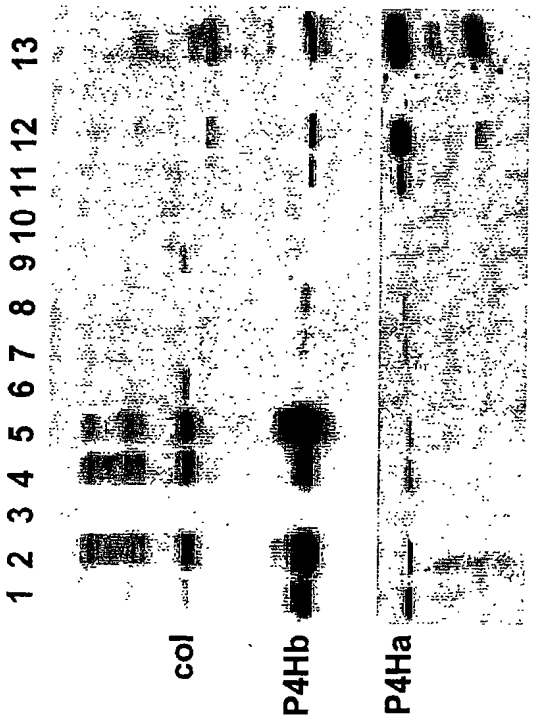


Figure 20

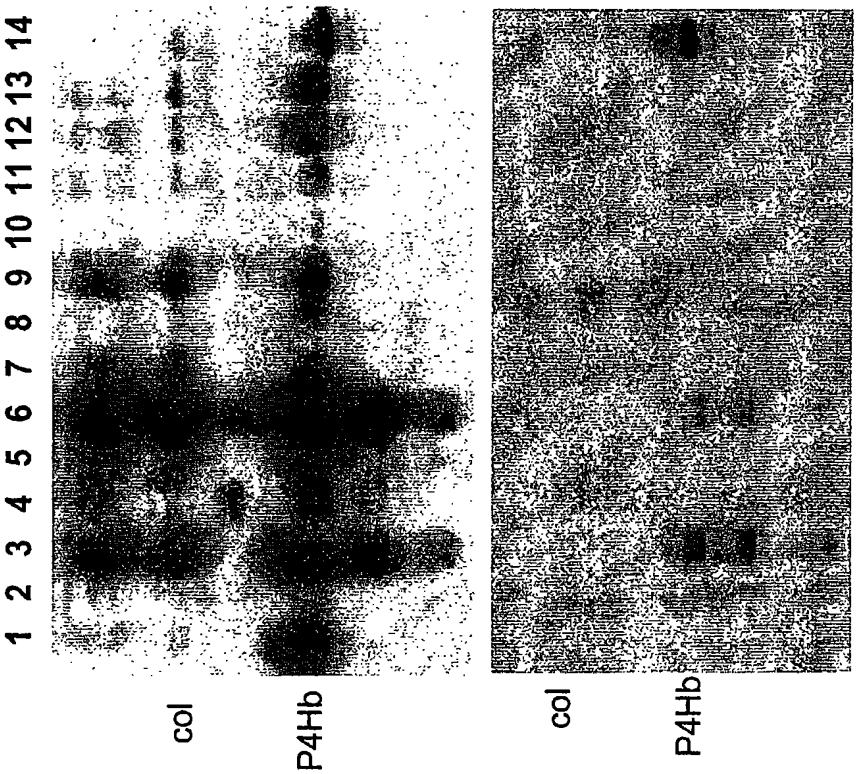


Figure 21

	Native signal peptide	Plant signal peptide
Collagen	...ILA//QQGEV...	...SSA//QQGEV...
P4Hα	...SLA//HPGFFT...	...SSA//MPGFFT...
P4Hβ	...VRA//DAPEE...	...SSA//MAPEE...

Figure 22

			pwrg4716 +	pwrg4716 +	pwrg4724 +
	pwrg4715	pwrg4723	pwrg4715	pwrg4723	pwrg4723
P4Hβ	na	na	16 (32)	16 (40)	21 (30)
P4Hα	na	na	3 (20)	8 (18)	9 (11)
Collagen	13 (36)	8 (36)	9 (20)	18 (19)	6 (11)
Triple Expressers	na	na	3 (32)	4 (18)	4 (11)

Figure 23

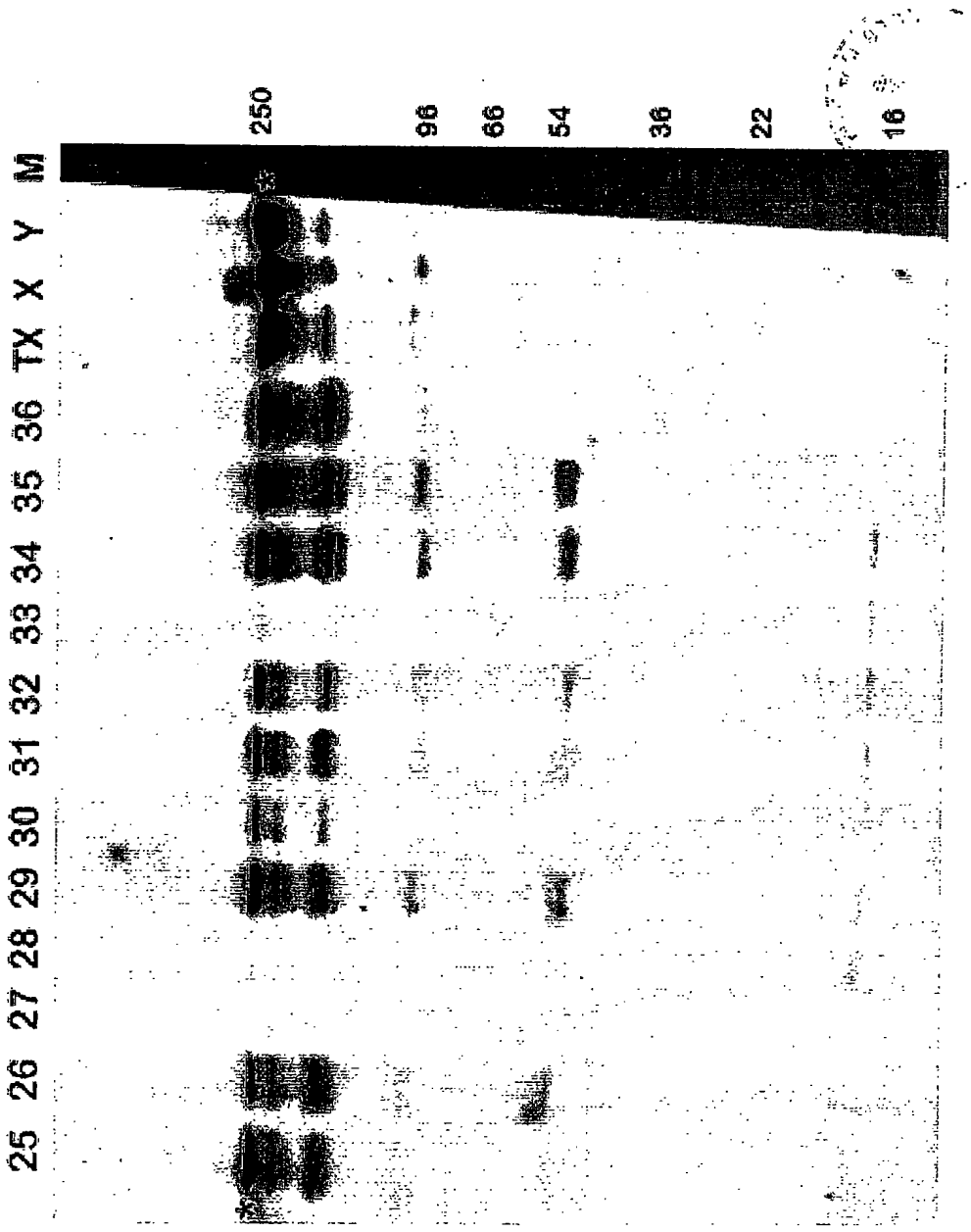


Figure 24